

Duquesne University Duquesne Scholarship Collection

Electronic Theses and Dissertations

Summer 2005

Arr Genes from Arsenate-Reducing Low G+C Gram Positive Bacteria *Bacillus selenitireducens* strain MLS10 and *Clostridium* sp. strain OhILAs

Mrunalini Ranganathan

Follow this and additional works at: <https://dsc.duq.edu/etd>

Recommended Citation

Ranganathan, M. (2005). Arr Genes from Arsenate-Reducing Low G+C Gram Positive Bacteria *Bacillus selenitireducens* strain MLS10 and *Clostridium* sp. strain OhILAs (Master's thesis, Duquesne University). Retrieved from <https://dsc.duq.edu/etd/1085>

This Immediate Access is brought to you for free and open access by Duquesne Scholarship Collection. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Duquesne Scholarship Collection. For more information, please contact phillips@duq.edu.

***arr* genes from arsenate-respiring low G+C Gram positive bacteria
Bacillus selenitireducens strain MLS10 and *Clostridium* sp. strain
OhILAs**

A Thesis Presented to the Bayer School of Natural and
Environmental Sciences
Department of Biological Sciences Duquesne University

By

Mrunalini Ranganathan

A thesis submitted in partial fulfillment of the
requirements for the degree of
Masters of Science

Thesis Advisor: Dr. John F. Stolz, Ph.D.

Thesis Committee:

Dr. Peter Castric Ph.D.

Dr. Nancy Trun Ph.D.

Submitted: July 11th 2005

Acknowledgements

I would like to thank Dr. Stolz for his enthusiasm, support and guidance during the past two years. His confidence in my abilities has taken me past several hurdles and through stressful times. I wish to thank Dr. Brady Porter for his timely suggestions and guidance and his entire lab, with special thanks to Christina Ventrice for her unending help with sequencing; Dr. David Lampe and the members in his lab for letting me use their spectrometer any time I wanted to and the McCormick lab for their generosity with lab supplies in cases of emergency.

Thanks are also due to my lab mates past and present, Miru for introducing me to the lab practices, Eddie for his accurate skills in media preparation, Brian for his enthusiasm and a never-ending helping hand, Lakshmi for her chirpy nature, Dustin and Robin for the extra hands and everyone who has worked in the Stolz lab. Dr. Mahmoud Berekaa will always remain a very good friend, philosopher and guide for his selfless manner and for teaching me to always believe in myself. Special thanks go to the faculty and staff for all their help.

Words alone cannot express the thanks I owe to Balaji, my husband, and my parents for giving me this opportunity for higher education and my sister and brother-in-law for their immense encouragement that has carried me smoothly through these two years. Last but not least, I will always remember Dr. Edward Weisberg for the support and encouragement he offered.

Abstract

The gene encoding the small subunit of the respiratory arsenate reductase (*arrB*) from the haloalkaliphilic low G+C Gram positive bacterium *Bacillus selenitireducens* strain MLS-10 was amplified by PCR, cloned and sequenced. The amplicon was larger (~1.2 kb) than predicted indicating that the degenerate reverse primer hybridized to a region further down stream. The amplicon contained all of *arrB* and the 5' end of another open reading frame. The *arrB* (693 bp) predicts a 26.3 kDa protein of 230 amino acids with a pI of 5.5 and four iron-sulfur binding domains. The protein shares a 50% identity and 66% similarity with ArrB from *Shewanella* sp. strain ANA-3. *arrA* and *arrB* were also successfully amplified from the freshwater low G+C Gram positive *Clostridium* sp. strain OhILAs using newly designed degenerate PCR primers. Interestingly, the *Clostridial* ArrA appears to be more closely related to *Chrysiogenes arsenatis* while its ArrB is more closely related to *Shewanella* sp. ANA-3. These results underscore the diversity in the *arr* operon.

Table of Contents

Title Page	
Signature Page	
Acknowledgements	i
Abstract	ii
Table of Contents	iii
List of Figures	v
List of Tables	xi
Abbreviation	xii
I. INTRODUCTION	1
Background	1
A. Arsenic in the Environment	2
1. Bangladesh- a tragic case	3
2. Arsenic contamination in the USA	5
B. Arsenic in Physiology	8
1. Toxicity of Arsenate and Arsenite	8
2. Metabolism	8
a. Arsenic Resistance for Detoxification	9
b. Microbial Arsenite Oxidation	11
c. Respiratory Arsenate Reduction	15
C. Genes involved in Respiratory Arsenate Reduction	18
1. ArrA – the catalytic subunit	19
2. ArrB – the Fe-Sulfur cluster containing subunit	21
3. Gene order	21
4. <i>Bacillus selenitireducens</i> – Physiology	22
5. Morphology of MLS10	23
II. RESEARCH GOALS	25
Specific aims	25
III. METHODS AND MATERIALS	26
Cell culture and DNA extraction	26
PCR and amplification with existing primers	28
Amplification of the Fe-S cluster subunit - <i>arrB</i> from MLS10	29
Purification of PCR product	30
Primer design	31
New primers for amplification of MLS10 <i>arrB</i> gene	32
Primers used for the amplification of <i>arrAB</i> gene from <i>Clostridium</i> sp. OhILAs	32
PCR optimization and amplification	33
Cloning of the PCR product	34
Plasmid purification and analyzing transformants	35
Sequencing from PCR product / plasmid DNA	36
Sequence analysis	39

IV. RESULTS	46
PCR and amplification with existing primers: MLS10 <i>arrA</i> amplification	46
Direct sequencing vs. cloning, transformation and sequencing	46
Amplification of the catalytic subunit <i>arrB</i> from MLS10	46
Direct sequencing from PCR product 34JSF3/ArrBDhR3 containing the <i>arrB</i> gene of interest	47
Sequencing the <i>arrB</i> gene from MLS10	54
Sequence analysis	56
Identification of the <i>arrB</i> gene	59
<i>Clostridium</i> sp. OhILAs – <i>arrA</i> and <i>arrB</i> genes	67
OhILAs <i>arrA</i> gene	71
Identification of the <i>arrB</i> gene in OhILAs	78
V. DISCUSSION	91
VI. FUTURE DIRECTIONS	95
VII. REFERENCES	96
VIII. APPENDICES	100
Appendix 1: BASAL SALT MEDIUM for <i>Bacillus selenitireducens</i> strain MLS10	100
Appendix 2: BASAL SALT MEDIUM for <i>Clostridium</i> sp. strain OhILAs	103
Appendix 3: Genomic DNA extraction protocol	104
Appendix 4: RNase treatment of genomic DNA protocol	105
Appendix 5: Direct sequencing protocol (Dr. David Lampe's Lab)	106
Appendix 6: Bacterial Media – Luria-Bertani (LB Medium), SOC and SOB Medium	108
Appendix 7: Protocol for Cracking Procedure	109

List of Figures

Figure 1: Arsenic concentrations in ground water in Bangladesh (British Geological Survey, 2000)	4
Figure 2: Top left, occurrence of arsenic (primarily from bedrock) in wells in New England (Ayotte et al., 2003). Bottom left, broiler poultry production in the US, 70% of which use organoarsenicals as a feed additive. Top right, arsenic in ground water in the US (Ryker, 2001). Bottom right, distribution of MSMA application.	7
Figure 3: Operon diversity of <i>aox</i> (arsenite oxidase) (Silver and Phung, 2005)	13
Figure 4: Phylogenetic diversity of representative respiratory arsenate-reducers (circles), heterotrophic arsenite-oxidizers (triangles), and chemolithautotrophic arsenite-oxidizers (boxes). (Oremland and Stolz, 2005)	14
Figure 5: Operon diversity in <i>arr</i> (respiratory arsenate reductase)	19
Figure 6: Whole mount (TEM) of a cluster of cells of <i>B.selenitireducens</i> strain MLS10 (dark spots indicate elemental selenium) (Oremland et al., 2004)	23
Figure 7: Position of primers 5-8 and 12-15. Clustal W(1.82) multiple sequence alignment of <i>arrB</i> gene from <i>Desulfotobacterium hafniense</i> , <i>Wolinella succinogenes</i> and <i>Shewanella</i> sp. ANA-3.	40
Figure 8: Position of primers 1-3, 9-11, 17-20 and 27-30. <i>Bacillus selenitireducens</i> strain MLS10 respiratory arsenate reductase catalytic subunit (<i>arrA</i>) gene – 2424 bases (AY283639.1) followed by an intergenic 29 bases and 87 bases of <i>arrB</i> gene 5' end.	43

Figure 9: Position of primers 21-26 and 31-35. Clustal W(1.82) multiple sequence alignment of *arrA* gene sequences from *Sulfurospirillum barnesii* SES-3, *Wolinella succinogenes*, *Bacillus selenitireducens* strain MLS10, *B.arseniciselenatis* strain E1H, *Desulfitobacterium hafniense* and *Shewanella* sp. strain ANA-3. _____ 45

Figure 10: Blast results for sequence obtained utilizing 80ng template DNA – 34JSF3/ArrBDhR3 and forward primer 34JSF3 where the query is the sequence supplied for the BLAST and the subject is the sequence/s that aligned the best from GenBank. (The stop codon for the *arrA* gene - TAA is marked in red. ATG is where *arrB* gene begins. New sequence is marked in blue. *arrA* and *arrB* are separated by an intergenic sequence of 29 bases) _____ 49

Figure 11: Blast results for sequence obtained utilizing 97ng template DNA – 34JSF3/ArrBDhR3 and forward primer 34JSF3 where the query is the sequence supplied for the BLAST and the subject is the sequence/s that aligned the best from GenBank. (The stop codon for the *arrA* gene - TAA is marked in red. ATG is where *arrB* gene begins. New sequence is marked in blue. *arrA* and *arrB* are separated by an intergenic sequence of 29 bases) _____ 51

Figure 12: *arrB* sequence obtained with 80ng DNA template (F3A), 97ng (F3B) & 48.5ng (F3C) followed by Clustal W (1.82) alignment of the 3 sequences (* indicates identical bases and the non-identical bases are shown green). _____ 52

Figure 13a: *arrX* from MLS10 (The forward primer MarrXF1 is shown in bold) _____ 57

Figure 13b: MLS10 Whole clone sequenced with the DArrBR3 reverse primer. (647 bases in blue correspond to *arrB* gene followed by 113 bases downstream) _____ 58

Figure 13c: MLS10 whole clone sequenced with DArrBR4 reverse primer. (524 bases in blue correspond to MLS10 *arrB* gene, 233 bases in black = MLS10 *arrA* gene 3') _____ 58

Figure 14: MLS10 ArrA protein sequence and conserved domain.

Above: the amino acid sequence of MLS10 ArrA- 807 residues, showing the cysteine-rich motif (C-X₂-C-X₃-C-X₂₇-C), the cysteine residues are shown **red**.

Below: Molybdopterin dinucleotide binding domain, residues 633-696. Identical residues are shown **red** and similar residues are shown **blue**._____ 60

Figure 15: MLS10 DNA and protein sequence (ArrB).

Above :The *arrB* gene –693 bases is shown in **blue** with **ATG** the start codon and **TAA** the stop codon in **bold**, RBS is underlined. Below: the amino acid sequence –230 residues corresponding to the above 693 bases excluding * - the stop codon. All the cysteine residues are shown in **red**._____ 61

Figure 16: Protein sequence alignment - ArrB from MLS10, *Wolinella succinogenes* and *Shewanella* sp. ANA-3 ([*] indicates amino acid identity, [. & :] indicate similarity, conserved cysteine residues are shown **red**)_____ 62

Figure 17: Neighbor joining, rooted tree with bootstrap values generated for protein sequences with significant alignment from the BLAST search for MLS10 ArrB.____ 63

Figure 18: Hydropathy profile of the MLS10 ArrB protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis._____ 65

Figure 19: OhILAs ArrA protein alignment ([*] indicates identity and [: & .] indicates similarity) _____ 71

Figure 20: Neighbor joining, rooted tree with bootstrap values generated using the protein sequences with significant alignment from the BLAST search for OhILAs ArrA._____ 73

Figure 21: Hydropathy profile of the OhILAs ArrA.1 protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis. _____75

Figure 22: Hydropathy profile of the OhILAs ArrA.2 protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis. _____75

Figure 23: Above :OhILAs *arrB* gene sequence 522 bases. Below: ArrB protein sequence 174 AA (* included). (*arrB* gene sequence is shown in blue, the RBS is underlined, the cysteine residues are in bold and red, the methionines in cyan, * indicates the stop codon for the ORF, inverted repeats are underlined and highlighted) _____78

Figure 24: OhILAs ArrB protein alignment with ArrB from MLS10, *Wolinella succinogenes* and *Shewanella* sp. ANA-3 ([*] indicates identity and [. & :] indicates similarity, all conserved cysteine residues are shown red) _____80

Figure 25: Rooted, neighbor joining tree with bootstrap values generated for protein sequences that showed significant alignments with OhILAs ArrB. _____81

Figure 26: Hydropathy profile of the OhILAs ArrB protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis. _____83

Figure 27: MLS10 *arrA* amplification; *arrB* amplification and cloning.

Panel A: MLS10 *arrA* amplification. Lane 1: MLS10 genomic DNA (~2-4 Mb). Lanes 2-4: PCR product amplified with primers 110JSF1/34JSR2 (~2.5 kb). **Panel B:** MLS10 *arrB* amplification. Lanes 1 and 2: PCR product amplified with primers 34JSF3/ArrBDhR3 (~2.5 kb) and 34JSF4/ArrBShR2 (~2.2 kb). **Panel C:** Cracking results of the cloned 2.5 kb amplicon containing MLS10 *arrB*. _____85

Figure 28: MLS10 *arrB* amplification and cloning.

Panel A: MLS10 *arrB* amplification. Lane 1: PCR product amplified with primers 34JSF4/ArrBDhR3 (~1.1 kb). **Panel B:** Cracking results of the cloned 1.1 kb amplicon containing MLS10 *arrB*. **Panel C:** Restriction digests with EcoR1 of selected clones. 86

Figure 29: Amplifying the *arrB* gene using newly designed primers.

Panel A: PCR product amplified with: primers MArrBF2/DArrBR3 (~1.4 kb) – Lane 1, primers MArrBF2/DArrBR1 (~556 bp) – Lanes 2 and 3, primers MArrBF3/DArrBR1 (~375 bp) – Lane 4.

Panel B: Lanes 1 and 2: PCR product amplified with primers MArrBF2/DArrBR3 (~1.4 kb) and MArrBF3/DArrBR3 (~1.2 kb), both dominant and expected fragments are seen, arrows indicate the dominant bands. Lane 3: PCR product amplified with primers MArrBF3/DArrBR4, arrows indicate two dominant bands. 87

Figure 30: PCR products amplified for sequencing MLS10 *arrB* gene.

Lane 1: PCR product amplified with primers MArrBF3/DArrBR2 (~520 bp) using a 1.4 kb fragment as DNA template amplified with primers MArrBF2/DArrBR3. Lane 2: PCR product amplified with primers MArrBF3/DArrBR2 (~520 bp). Lanes 3 and 4: PCR product (~1.4 kb, MArrBF2/DArrBR3) and Lanes 5 and 6: PCR product (~1.2 kb, MArrBF3/DArrBR3) amplified for elution and purification. 88

Figure 31: OhILAs *arrA* gene amplification and sequencing.

Panel A: PCR product amplified with: primers arrAUF1/arrAUR2 (~245 bp) – Lanes 1 and 3, primers arrAUF1/arrAUR3 (~852 bp) – Lane 2, primers arrAMLSF4/arrAUR6 (~348 bp) – Lane 4.

Panel B: PCR products amplified for sequencing OhILAs *arrA* gene. Lanes 1, 4 and 5: PCR product amplified with primers arrAUF1/arrAUR3 (~852 bp). Lanes 2, 6 and 7: PCR product amplified with primers arrAMLSF4/arrAUR6 (~348 bp). Lanes 3, 8 and 9: PCR product amplified with primers arrAUF1/arrAUR5 (~1825 bp). 89

Figure 32: OhILAs *arrB* gene amplification and sequencing.

Lanes 1,4 and 5: PCR product amplified with primers MArrBF3/DArrBR1 (~375 bp).

Lane 2: PCR product amplified with primers arrAUF5/DArrBR4 (~800 bp). Lane 3: PCR product amplified with primers arrAUF5/DArrBR2 (~1.2 kb)._____ 90

List of tables

Table 1: Primers used in this study. * 34JSF4 was wrongly designed based on vector sequence. _____	41
Table 2: Primer pairs used for amplification of <i>arrB</i> from MLS10, T _m values, expected fragment size and the optimal annealing temperature for PCR. Corrected fragment sizes are in parentheses, ____ indicates no optimal annealing temperature. _____	56
Table 3: Amino acid composition of the predicted product of the <i>arrB</i> gene from MLS10 _____	66
Table 4: Primer pairs used for amplification of <i>arrA</i> and <i>arrB</i> from OhILAs, T _m values, expected fragment size and optimal annealing temperature for PCR. _____	67
Table 5: Amino acid composition of the predicted product of the +2 ORF (152 –988) from OhILAs corresponding to the first half of the <i>arrA</i> gene (<i>arrA.1</i>). _____	74
Table 6: Amino acid composition of the predicted product of the +1 ORF (1075 –2385) from OhILAs corresponding to the second half of the <i>arrA</i> gene (<i>arrA.2</i>). _____	76
Table 7: Amino acid composition of the predicted product of OhILAs <i>arrB</i> gene. ____	82

ABBREVIATION

Arr – respiratory arsenate reductase

Aox – Arsenite oxidase

Aso – Arsenite oxidase

Ars – Arsenate resistance system

ARMs – arsenate-resistant microorganisms

DARPs – dissimilatory arsenate-reducing prokaryotes

CAOs – chemoautotrophic arsenite oxidizers

HAOs – heterotrophic arsenite oxidizers

DMSO – Dimethyl sulfoxide

MSMA – Mono sodium methyl arsonate

DSMA – Disodium methyl arsonate

ORFs – Open Reading Frames

RBS – Ribosome Binding Site

SDS – Sodium Dodecyl Sulfate

I. INTRODUCTION

Background

Arsenic (As), a trace element with a crustal abundance of 0.0001% is only the 20th most common element in the earth's crust. It is naturally present in air, soil, rocks and water. Arsenic has a rich history behind it, with its extensive use in the Fowler's solution as a tonic and towards treatment for psoriasis and asthma. Arsenic was extensively used in the manufacture of pesticides and fungicides until very recently, when these were largely replaced by chlorinated hydrocarbons (Hindmarsh, 2000). The most well known use of arsenic trioxide (As_2O_3) is as a homicidal poison in the medieval times referred to as "inheritance powder". It is still used today for more beneficial ends as a treatment of certain forms of leukemia. The industrial use of arsenic still continues in the production of pigments, semiconductors etc. In agriculture, organic forms are used as desiccants and defoliants and in swine and poultry feed (Roxarsone – 3 nitro-4-hydroxyphenylarsonic acid) (Oremland and Stolz, 2003).

In the environment, arsenic can exist in any one of these four oxidation states: arsine [$\text{As}(-\text{III})$], metallic [$\text{As}(0)$], arsenate [$\text{As}(\text{v})$] and arsenite [$\text{As}(\text{III})$]. Arsenate the more oxidized species is a common inorganic form in aqueous aerobic environments whereas arsenite the more toxic species is prevalent in anoxic environments. One rarely comes across native [$\text{As}(0)$] and traces of gaseous arsine.

A. Arsenic in the Environment

Arsenic enters the environment ultimately by weathering of rocks and arsenic containing minerals, volcanic eruptions and emissions and other natural phenomenon that release different arsenic species into the soil, water or air. The anthropogenic release of arsenic, however, is an alarming 28,000 metric tons per year, nearly four times the natural input (Cullen, 1989.) Arsenate toxicity can affect the skin, several vital organs including the liver, kidneys, lungs and prostate glands, can cause diabetes, peripheral neuropathy and coronary disease (Hindmarsh, 2000). Anthropogenic activities including the burning of fossil fuels, smelting of ore and mining have elevated the levels of arsenic in several terrestrial and aquatic environments. Arsenic poses a threat to wildlife and human health due to its occurrence at increasingly high concentrations in these environments. Arsenic concentration in soil samples from Northern Chile has been reported to be as high as 1099 mg kg^{-1} . Aquifers also have been found to contain 11.25 mg L^{-1} of arsenic. In this case geological factors have contributed to this increase in arsenic content.

Arsenic is a public health hazard in many parts of the world, including Bangladesh, West Bengal (India), Mongolia, Taiwan, and Vietnam; drinking water being the most significant source of arsenic (Nriagu, 1994). This high arsenic concentration in drinking water maybe attributed to several different facts. For example, natural hot springs can contain anywhere from 1 to 10mg/L of arsenic, whereas sites with biological activity where microbes are engaged at transforming the sediments resulting in the speciation and mobilization of arsenic are the other possibility that has brought a lot of excitement into this subject. Arsenic concentrations up to 50 mg L^{-1} have been reported in some areas where biological activity is speculated to be the cause for the such phenomenon (Gihring

and Banfield, 2001). Hence, knowledge about the biological mechanisms of microbial arsenic transformation is becoming very important (Martin et al., 2001). Removal of arsenic from the drinking water is a primary environmental concern.

1. Bangladesh, a tragic case

In recent years, however, the intervention of humans has led to the high frequency of arsenicosis (a chronic illness resulting from drinking water with high levels of arsenic over a long period of time; it is also known as arsenic poisoning), which has put millions of people at risk in the Ganges delta of Bangladesh and West Bengal (India). The drinking water from shallow wells has levels of arsenic significantly above the acceptable limit of 50 ppb set for Bangladesh (Figure 1). These wells were dug over the last two decades, by the government of Bangladesh with help from the UNO, which suggested this as an alternative for safe drinking water until the catastrophe came in the form of arsenic contamination. It seems likely that microbial metabolism is involved in mobilizing arsenic rich layers of the Pleistocene in these shallow wells of Bangladesh (Silver and Phung, 2005).

Arsenic is a major constituent of more than 200 minerals most of which are in the form of ores, in combination with other elements. Arsenopyrite, realgar, orpiment and scorodite are some arsenic containing minerals. Arsenate has a high affinity for insoluble compounds like ferrihydrite and alumina and is readily adsorbed into the surface of such minerals in sediments and rocks. Arsenite on the other hand, being less reactive, becomes the more mobile oxyanion of the two. In the case of Bangladesh, iron containing sediments are significant in that the biogeochemical cycling of iron and

arsenic are linked. It has been demonstrated earlier that dissimilatory arsenate reducing prokaryotes (DARPs) like *Sulfurospirillum arsenophilum*, release As(III) from an initial solid phase consisting of ferrous arsenate equivalent to the case of the aquifers in Bangladesh, where the toxic As(III) is now mobilized from the sediments into the water, the source of drinking water. The problem does not end at this stage, but goes further with our knowledge of microbes that can oxidize arsenite, reduce and respire arsenate (Mukhopadhyay et al., 1998; Oremland and Stolz, 2003; Croal et al., 2004; Oremland and Stolz, 2005).

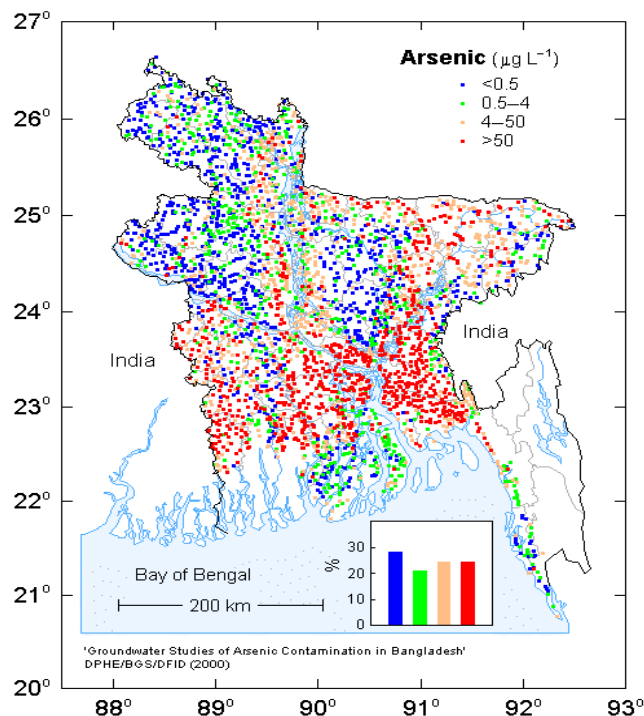


Figure 1: Arsenic concentrations in ground water in Bangladesh (British Geological Survey, 2000)

2. Arsenic contamination in the USA

Around 50% of drinking water in the United States comes from groundwater. Arsenic contamination is a public health concern because a recent study of 1,600 public and private water sources in six New England states indicated that close to a third had levels of arsenic exceeding 10 ppb (Ayotte et al., 2003)(Figure 2,Top left).

The World Health Organization has set the acceptable level for arsenic in drinking water at 0.01 mg/L (10 ppb) and the United States legislature has brought down the earlier five fold higher level of 50ug/L to that set by the WHO. The U.S. Environmental Protection Agency has also recently adopted this standard. Regions throughout the United States including the mid and southwest have elevated levels of arsenic in water as per a survey by The USGS NAWQA (Figure 2, Top right).

The western United States including western Nevada and eastern California is a volcanic area. Several alkaline lakes in this region have naturally occurring high concentrations of dissolved inorganic arsenic ranging from 3000uM in the Searles Lake to 0.8uM in the Crowley Lake. Mono Lake has As concentrations lying in-between the two extremes (200uM) which is still on the higher side where arsenic is readily cycled and can be important in carbon mineralization (Oremland et al., 2004).

The use of inorganic arsenic in manufacture of pesticides has been banned and the use of CCA in wood preservatives voluntarily discontinued. But organic arsenicals like roxarsone and monosodium and disodium methyl arsonate (MSMA and DSMA) are still widely used (Figure 2, Bottom left). Roxarsone is used as a feed additive for poultry to prevent coccidiosis and thereby increase the appetite and influence growth and pigmentation of the chickens. It has been shown that roxarsone is degraded with the

involvement of biological processes but neither the microorganisms involved nor the products produced have been identified (Garbarino et al., 2003). It is estimated that 70% of the broiler chickens raised in the United States are fed roxarsone and 20-50 metric tons are used annually on the east coast alone (Christen, 2001; Garbarino et al., 2003) (Figure 2, Bottom left) MSMA and DSMA are used in the production of cotton and for the maintenance of golf courses. It has been estimated that in 1997, 2,434 tons of MSMA and 421 tons of DSMA were applied to cotton, citrus, and sod (Figure 2, Bottom right).

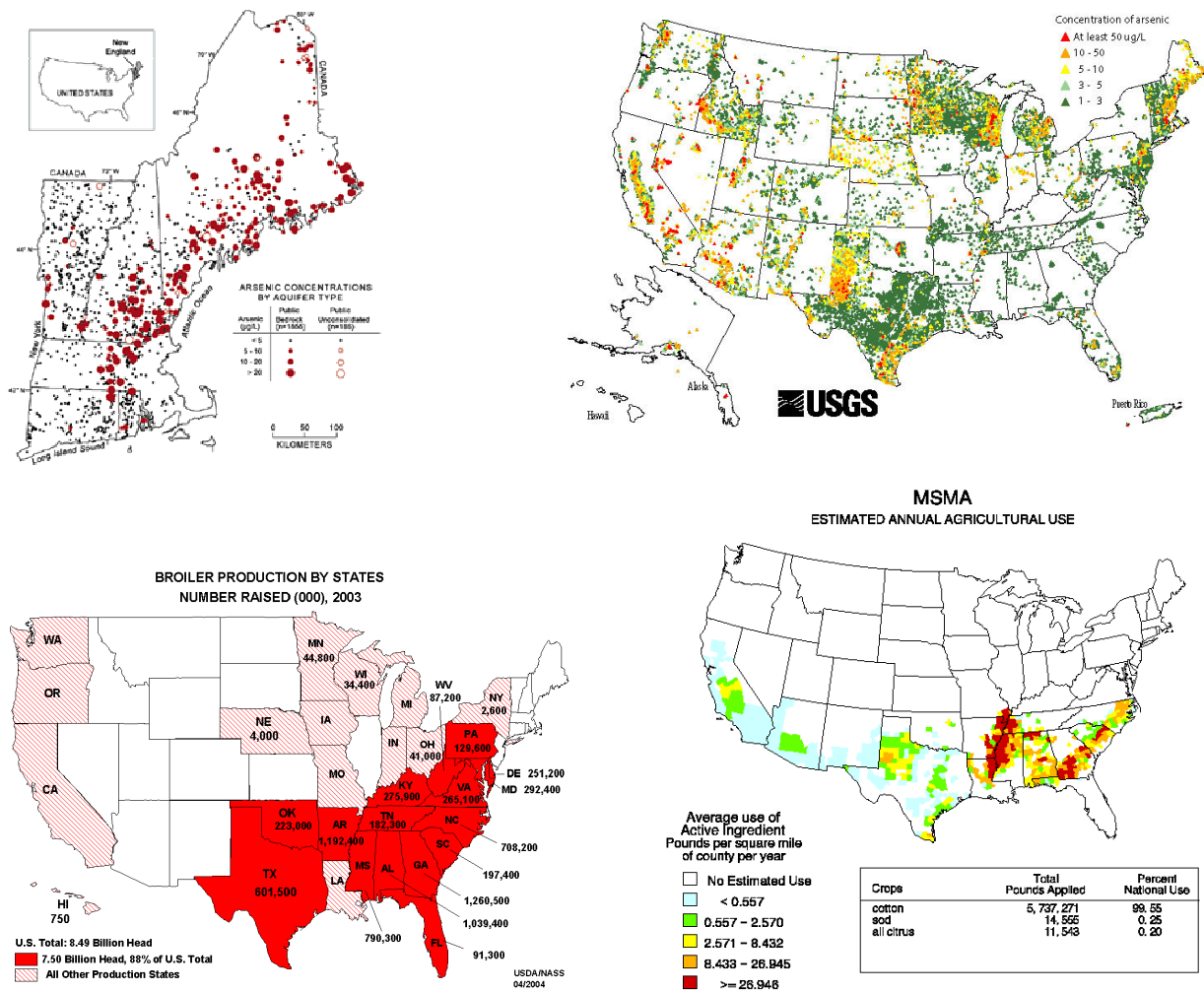


Figure 2: Top left, occurrence of arsenic (primarily from bedrock) in wells in New England (Ayotte et al., 2003). Bottom left, broiler poultry production in the US, 70% of which use organoarsenicals as a feed additive. Top right, arsenic in ground water in the US (Ryker, 2001). Bottom right, distribution of MSMA application.

B. Arsenic in Physiology

1 . Toxicity of As(V) and As(III)

The two inorganic forms arsenate and arsenite are toxic due to their chemical and molecular forms. Arsenate (HAsO_4^{2-} and $\text{H}_2\text{AsO}_4^{1-}$) is a molecular analogue of phosphate and inhibits oxidative phosphorylation, thus disrupting respiration, the energy-generating source in humans. Arsenite [H_2AsO_3^-] on the other hand, with its capacity to readily bind to reactive thiol groups (SH) of several enzymes, acts by destabilizing these enzymes (Oremland and Stolz, 2003; Silver and Phung, 2005). Acute and chronic exposure to these forms, have severe carcinogenic effects ranging anywhere from the “Blackfoot disease” in Taiwan to coronary artery disease and carcinoma of the liver, kidney, bladder, lung and the most common of them all, carcinoma of the skin in Bangladesh (Hindmarsh, 2000; Anawar et al., 2002).

2. Metabolism

Microbes have evolved mechanisms to cope with high levels of arsenic species in their environment. They have in their repertoire mechanisms for arsenic resistance and enzymes that oxidize As(III) to As(V) or reduce As(V) to As(III). The primordial atmosphere being a reducing environment maintained high levels of arsenite, and therefore, arsenite efflux pumps would have sufficed for arsenic detoxification. With the introduction of oxygen into the atmosphere by cyanobacteria 3.8 billion years ago, arsenate, the more oxidized species was encountered more often. This provided a selection pressure for the convergent evolution of three different arsenate reductases,

which would work in conjunction with the already existing efflux pump (Mukhopadhyay and Rosen, 2002).

a. Arsenic Resistance for detoxification

The ArsC system confers resistance in arsenate resistant microbes (ARMs) . There is no gain of energy through this process and is a mechanism to cope with high levels of arsenic in the environment. The proteins that provide the mechanism of arsenic resistance are encoded by the *ars* operons. In bacteria, heavy metal resistance genes are frequently located on plasmids, but chromosomal locations are also known (Cervantes et al., 1994).

The arsenate reductase (ArsC) enzyme is found in Gram-positive and Gram-negative bacteria as well as in the eukaryotic yeast *Saccharomyces cerevisiae*. Recent studies based on evolutionary trees of protein sequence similarities show three separate clades of arsenate reductases, wherein the proteins share a common biochemical function but have no evolutionary relationship (Mukhopadhyay et al., 2002).

The first and best studied arsenate reductase is that of the Gram-negative bacterium *Escherichia coli* . The *ars* operon confers resistance to arsenic and antimony in *E.coli* (Carlin et al., 1995). The plasmid (R773 plasmid) encoded operon consists of the *arsR*, *arsD*, *arsA*, *arsB* and *arsC* genes whereas the chromosomal encoded *ars* has only *arsR*, *arsB* and *arsC* (Martin et al., 2001). ArsC is a small protein (13 – 15 kDa) that mediates the reduction of As(V) to As(III), in the cytoplasm. Arsenite, the more toxic form, can then be expelled out of the cell by ArsB - an As(III) specific transporter. ArsA is a soluble ATPase subunit activated by the presence of heavy metals. It forms a complex with ArsB, forming an ATP dependent As(III) transporter –ArsAB. The

regulatory proteins, ArsR and ArsD are involved in the expression of the *ars* operon (Gladysheva et al., 1994; Rosen, 2002).

Three cysteines are used for the reduction of As(V) to As(III). A cysteine residue (Cys12) near the N-terminal of ArsC binds the arsenate and the other two come from reduced glutaredoxin and glutathione, the reducing power for the reaction. Similarly, six amino acids including the Cys12 have been identified from studies on the plasmid R773 that are completely conserved within the first clade (His8, Ser12, Arg60, Arg94 and Arg107) (Mukhopadhyay et al., 2002).

Plasmid pI258 of *Staphylococcus aureus* is representative of Gram-positive plasmid encoded systems. Its *ars* operon consists of only *arsB*, *arsC* and *arsD*. Reduced thioredoxin provides the electrons to reduce As(V) to As(III) via the ArsC. Subsequently, the As(III) is expelled out of the cell by the ATP independent ArsB. The plasmid pI258 operon shares low homology (10%) with the plasmid R773 encoded system of *E.coli* but uses a similar cysteine residue. The pI258 proteins are similar to the family of low molecular weight protein phosphotyrosine phosphatases found in microbes and animals, including humans (Bennett et al., 2001). Although arsenic resistance and detoxification has been studied in detail in *E. coli* and *S. aureus*, several other bacteria including strict anaerobes like *Clostridium* and *Desulfovibrio* (Krafft et al., 2000) as well as aerobes isolated from As-contaminated soils and mine tailings are reported to possess this mechanism, thereby suggesting that the geochemical cycling of arsenic depends on such processes.

The yeast *Saccharomyces cerevisiae* represents a third type of *ars* operon. In yeast, the *ars* operon involves three genes - *acr1*, *acr2*, and *acr3*. Mutants of *acr1*

mutants are hypersensitive to arsenate and arsenite whereas deletion of *acr3* renders the cell incapable of resistance to both oxyanions while *acr2* mutant is resistant to arsenite alone (Mukhopadhyay and Rosen, 1998). The regeneration of the active Acr2p is achieved by the reduction of the disulfide bond with glutathione by glutaredoxin.

b. Microbial Arsenite Oxidation

Bacterial arsenite oxidation was first reported in 1918 and since then organisms representing over 30 strains of at least 9 genera including α , β , and γ proteobacteria, Deinocci (Thermus) and Crenarcheota have been found. These include the Deinocci-*Thermus aquaticus* and *Thermus thermophilus*, that have been found to rapidly oxidize As(III) to As(V) in their natural environments of terrestrial hot springs or otherwise thermally polluted water systems (Gihring et al., 2001). *Alcaligenes faecalis* and five members of the β proteobacteria are heterotrophic arsenite oxidizers (HAOs) whereas *Pseudomonas arsenitoxidans* and “NT-26” are chemolithautotrophic arsenite oxidizers (CAOs) (Figure 4) (Oremland and Stolz, 2005). CAOs are bacteria that oxidize As(III) [H_2AsO_3^-] to As(V) [HAsO_4^{2-}] utilizing oxygen or nitrate as the terminal electron acceptor ultimately for the fixation of inorganic carbon (CO_2). The HAOs on the other hand respire oxygen but cannot grow without organic carbon as a source of energy. Unlike the very dissimilar arsenate reductases (ArsC), both the classes of arsenite oxidizers have notable similarities in their oxidases (Aox).

Six members of α Proteobacteria (Ben-5, NT-3, NT-4, NT-2, NT-26 and NT-25) and one member of γ Proteobacteria (MLHE-1) are known chemolithautotrophic arsenite oxidizers (Figure 4). Strain MLHE-1, a γ proteobacterium isolated from Mono Lake uses nitrate as the terminal electron acceptor (Oremland et al., 2002).

The most well studied of all the arsenite oxidizers is *Alcaligenes faecalis*, a heterotrophic arsenite oxidizer (Osborne and Enrlich, 1976). The arsenite oxidase from this proteobacterium has been purified and characterized with crystal structure analysis of the enzyme placing it in the DMSO reductase family of molybdoenzymes similar to the respiratory arsenate reductases (Arr) (Anderson et al., 1992).

The arsenite oxidase in *A. faecalis* involves a periplasmic electron transfer chain consisting of an arsenite oxidase, azurin, cytochrome C and cytochrome C oxidase. Aerobic growth on arsenite is favorable wherein the reaction is exergonic, whereas, respiration on the same substrate is not acceptable, the latter being a means of detoxification with no release of energy.

The arsenite oxidase has 2 subunits; a 88 kDa subunit encoded by the *asoA* gene containing the molybdenum in conjunction with two pterin cofactors and a smaller 14 kDa subunit encoded by the *asoB* gene with a Rieske 2Fe-2S cluster motif (Anderson et al., 1992) that is typical of arsenite oxidizers. A significant difference between the Aso and Arr, both belonging to the DMSO reductase family of molybdoenzymes is the molybdenum linkage to the enzymes. The molybdenum in AsoA is not directly linked by covalent bonds to the side chains of the conserved cysteine, selenocysteine, alanine or serine residue in the protein as in the dimethyl sulfoxide (DMSO) reductases; there is no covalent bond between the Mo atom and the protein (Ellis et al., 2001).

The arsenite oxidase from NT-26 was fully characterized at the molecular and biochemical level. The periplasmic enzyme again has two subunits like the Aso from *A. faecalis*; AroA – a 98 kDa subunit, an enzyme with a characteristic conserved cysteine motif (C-X₂-C-X₃-C) similar to that found in ArrA (C-X₂-C-X₃-C-X₂₇-C), the larger

subunit of the arsenate reductase heterodimeric proteins, which places the enzyme in the DMSO reductase family of molybdoenzymes and a smaller subunit AroB –a 14 kDa subunit, the first periplasmic protein to be identified with a Rieske 2Fe-2S cluster motif (vanden Hoven and Santini, 2004).

Aro and Aso are very similar but for the difference in the subunits where Aro has an $\alpha_2\beta_2$ subunit configuration and Aso has only an $\alpha_1\beta_1$ configuration of the subunits (Figure 3). Interestingly, NT-26 is found to be resistant to high concentrations of arsenate (100mM) probably conferred by a homologous *arsC* gene it contains (Figure 3). Its involvement in energy generation is another possibility wherein mutants for the oxidase are not capable of chemolithautotrophic growth unlike the case of *A. faecalis* whose mutants for the *aso* gene showed no difference in growth (vanden Hoven and Santini, 2004).

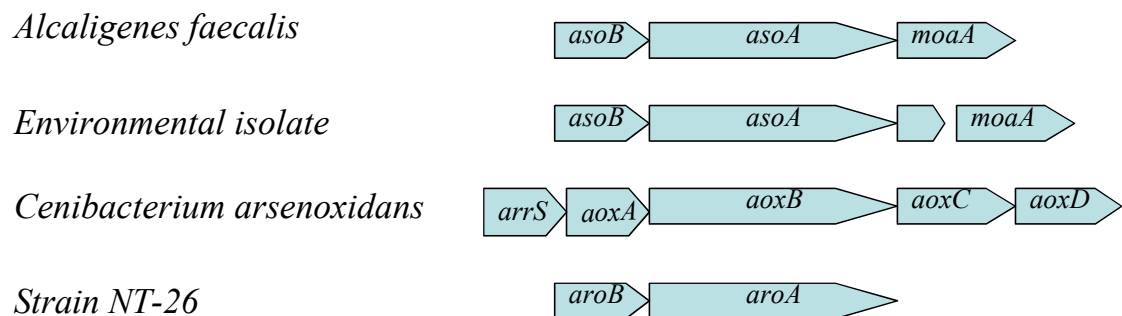


Figure 3: Operon diversity of *aox* (arsenite oxidase) (Silver and Phung, 2005)

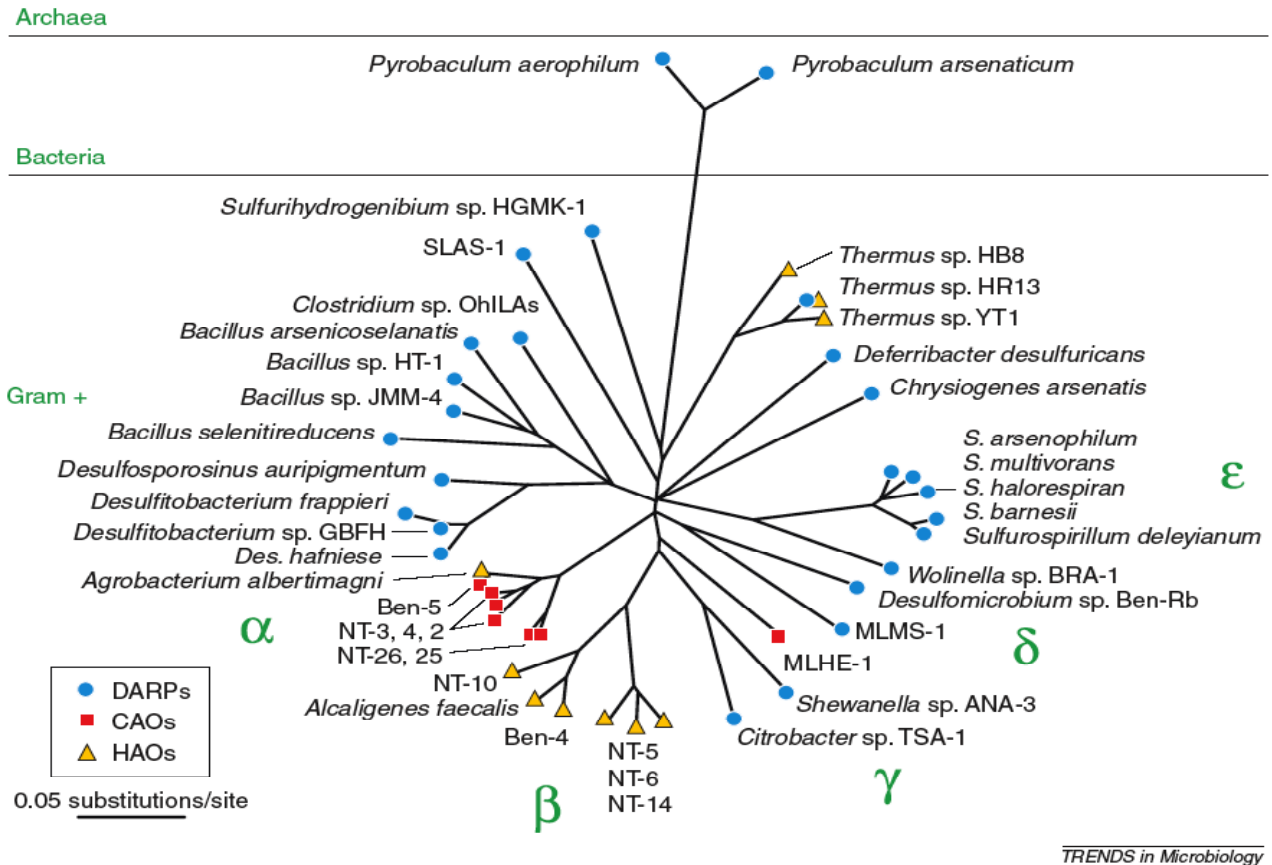


Figure 4: Phylogenetic diversity of representative respiratory arsenate-reducers (circles), heterotrophic arsenite-oxidizers (triangles), and chemolithautotrophic arsenite-oxidizers (boxes). (Oremland and Stolz, 2005)

Thermus HR13 is an interesting member of this class of arsenite oxidizers (Figure 4), because, it has been shown to use As(V) for respiration in oxygen deficient environments such as the Growler Hot Springs. This facultative anaerobe when grown with arsenate as the sole electron acceptor doubles in cell mass (Gihring and Banfield, 2001).

c. Respiratory Arsenate Reduction

The discovery that arsenic could be used as a life supporting substrate came as a big surprise to many in the scientific world (Ahmann et al., 1994). The reduction of arsenate to arsenite for detoxification purposes is important for the biogeochemical cycling of the toxic element. It is now seen that bacteria use an arsenate reductase to reduce arsenate to arsenite coupled to the oxidation of organic substrates or hydrogen ultimately resulting in energy production.

The As(V) /As(III) oxidation reduction potential is +135mV, which is thermodynamically favorable when coupled to oxidation of organic matter . To date, 27 dissimilatory arsenate-reducing prokaryotes are known. These are anaerobes from the domain bacteria and archaea including 11 Gram-positive bacteria, 8 ϵ -Proteobacteria, 3 δ -Proteobacteria and 2 Archaea (Oremland and Stolz, 2005) (Figure 4). At least 16 species of arsenic respiring bacteria have been isolated in pure culture, consisting of representatives from the above domains (Oremland and Stolz, 2003).

Sulfurospirillum arsenophilum and *S. barnesii*, two closely related members of the ϵ – Proteobacteria were the first microbes found to reduce arsenate for dissimilatory purposes by coupling it with the oxidation of lactate (Stolz et al., 1999). The isolate *S. arsenophilum* MIT-13 was obtained from arsenic contaminated sediments of the Aberjona watershed, MA, USA (Ahmann et al., 1994). This bacterium is able to reduce arsenate to arsenite by oxidizing lactate to CO₂ at the same time, thereby driving growth, and producing energy needed for the same. It can also use H₂ + acetate, pyruvate and fumarate as alternative electron donors. Similarly electron acceptors also include nitrate and fumarate. *S. barnesii* strain SES-3 was isolated from a selenate rich marsh in

Western Nevada. SES-3, like MIT-13 can use a variety of electron donors such as acetate, pyruvate, lactate and H₂ and electron acceptors including selenate, nitrate, nitrite, fumarate, Fe(III), thiosulfate, elemental sulfur, DMSO and trimethylamine oxide (Stolz and Oremland, 1999).

The first sulfate reducing bacterium shown to grow with As(V) as the terminal electron acceptor was a Gram-positive bacterium, *Desulfosporum auripigmentum* strain OREX-4, a freshwater sediment isolate from the Upper Mystic Lake in Winchester, MA, USA (Newman et al., 1997). It metabolizes compounds used by sulfate reducers such as H₂, H₂ + acetate, butyrate, pyruvate, lactate, malate, glycerol and ethanol while using sulfate, sulfite, thiosulfate, As(V) and fumarate as electron acceptors. This was an added fact to the phylogenetic and physiological diversity of dissimilatory arsenate reducing prokaryotes (DARPs).

The first arsenate reductase (Arr) to be purified and characterized was that from *Chrysiogenes arsenatis*, and it differs from the detoxifying arsenate reductases ArsC of *E.coli* or *S.aureus*. The respiratory arsenate reductase is a soluble periplasmic enzyme comprised of heterodimeric subunits – ArrA (87 kDa) and ArrB (29 kDa). Molybdenum, iron –sulfur clusters and zinc are used as cofactors. The enzyme is As(V) specific and is induced only when As(V) is available as the electron acceptor (Macy et al., 1996). Thus, a periplasmic location, presence of cofactors, heterodimeric composition of the catalytic subunits and an ultimate gain of energy from the reaction, place the arsenate reductases from the detoxifying microorganisms (ArsC) very separate from the arsenate reductases (Arr) of the DARPs.

The diversity of the DARPs was further established with the isolation of two *bacillus* species from Mono lake, California, an alkaline, hypersaline, arsenic rich water body (Switzer Blum et al., 1998); basically an extreme environment different from all the above mentioned isolation sites. *B.arseniciselenatis* strain E1H and *B.selenitireducens* strain MLS10 are members of low G+C Gram-positive bacteria capable of reducing Se(VI) to Se(IV) and Se(IV) to Se(0) respectively, while utilizing a number of electron donors, completing an entire cycle for the geochemical cycling of selenium or in other words the speciation of the metal selenium in their natural environment (i.e. Mono lake).

The respiratory arsenate reductase of MLS10 is a membrane bound heterodimer (150 kDa) composed of a single α ArrA (110 kDa; *arrA* = 2424 bp) (Afkar et al., 2003) and β ArrB (34 kDa; *arrB* = 693 bp) (This study) subunit. The apparent k_m for arsenate is 34 μ M and the V_{max} is 2.5 μ mol min⁻¹ mg⁻¹. It is a member of the DMSO reductase family of molybdopterin. Optimal growth at pH 9.5 and 150 g/l NaCl indicate the topology of the enzyme to be oriented on the side of the cytoplasmic membrane that faces the peptidoglycan layer of cell wall similar to Gram-negative arsenate respiring bacteria (Afkar et al., 2003).

The arsenate reductase enzyme cannot reduce fumarate, nitrate or nitrite. Thus, additional terminal reductases are used for these respiratory pathways. The ArrA subunit contains the molybdenum catalytic center and ArrB is the iron-sulfur cluster containing subunit. Unlike the arsenate reductase from *C.arsenatis*, the enzyme from *B.selenitireducens* is localized to the membrane fraction (Afkar et al., 2003).

C. Genes involved in Respiratory Arsenate Reduction

Anaerobic respiration of arsenate is carried out by a phylogenetically diverse and metabolically versatile group of organisms. They are members belonging to the phylogenetic groups γ , δ , ϵ - Proteobacteria, low G+C Gram-positive bacteria and Archaea. There are at least 16 strains isolated in pure cultures (Oremland and Stolz, 2003; Oremland and Stolz, 2005). However, only the Arr's from *Chrysiogenes arsenatis* and *B. selenitireducens* have been purified and characterized (Krafft and Macy, 1998; Afkar et al., 2003) and genetically identified in *Shewanella* sp. strain ANA-3 (Saltikov and Newman, 2003). Based on these results, homologs of the arsenate reductases have been identified in the genomes of *Desulfitobacterium hafniense* and *Wolinella succinogenes*.

The arsenate reductase genes are encoded by the 'arr' operon. *arrA* encodes the large catalytic subunit, and *arrB*, the smaller electron transfer protein. Other genes may be involved in arsenate respiration but it has been shown that *Shewanella* sp. ANA-3 with only two open reading frames - *arrAB* operon (Saltikov and Newman, 2003) is able to respire arsenate. It does not require the *arsC* for As(V) respiration, showing clearly that the *arrA* and *arrB* genes are sufficient for anaerobic arsenate respiration (Saltikov et al., 2003). A gene found immediately upstream of the *arrA* gene in *W.succinogenes* and *D.hafniense*, *arrC*, encodes a membrane integral protein similar to NrfD (the membrane anchoring polypeptide of the dissimilatory pentaheme cytochrome c nitrite reductase). It has been hypothesized that ArrC is the anchoring protein for the Arr complex. Several

other genes are also thought to be a part of the *arr* operon, such as *arrD* in *D. hafniense*, an open reading frame immediately downstream of *arrB*, that is predicted to encode a chaperon protein similar to TorD of TMAO reductase. In addition, the genes for a two component regulatory system (*arrS*, *arrK*, *arrR*), also lie immediately upstream of *arrC* (Figure 5).

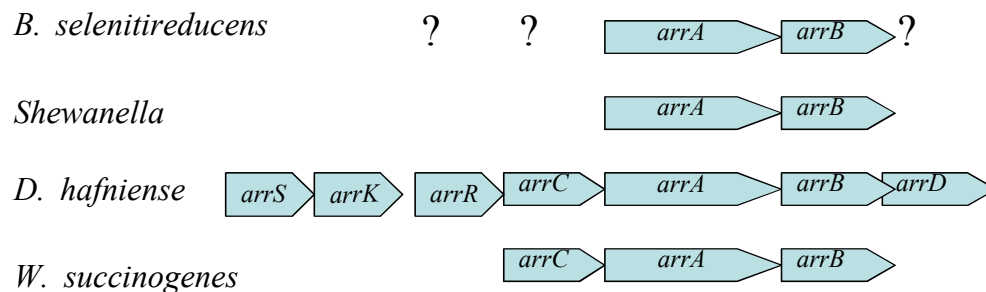


Figure 5: Operon diversity in *arr* (respiratory arsenate reductase)

The arsenate reductase operon *arr* is yet to be studied in detail in order to assign the genes involved as essential and/or necessary. Whole genome sequencing has been undertaken for four arsenate respiring bacteria, two belonging to low G+C Gram-positive group – *B.selenitireducens* strain MLS10 and *Clostridium* sp. Strain OhILAs; MLMS-1 a δ –Proteobacterium and *S.barnesii* a ϵ –Proteobacteria as part of the Joint Genome Project, Microbial Genome Project, Department Of Energy (DOE), USA.

1. *ArrA*- the catalytic subunit

ArrA is the larger subunit of the arsenate reductase. It is the molybdenum containing catalytic subunit. The ArrA protein has motifs, which are typical to molybdoenzymes belonging to the DMSO reductases family. The N-terminal sequence has a cysteine rich

motif (C-X₂-C-X₃-C-X₂₇-C) that is predicted to co-ordinate an iron-sulfur cluster and molydopterin dinucleotide binding domains one near the N-terminus that includes the putative ligand to the molybdenum and another at the C-terminus. Molydopterin dinucleotide binding domain is found in various molydopterin - containing oxidoreductases and tungsten formylmethanofuran dehydrogenase subunit d (FwdD) and molybdenum formylmethanofuran dehydrogenase subunit (FmdD); where the domain constitutes almost the entire subunit. The formylmethanofuran dehydrogenase catalyses the first step in methane formation from CO₂ in methanogenic archaea and has a molydopterin dinucleotide cofactor. This domain corresponds to the C-terminal domain IV in dimethyl sulfoxide (DMSO) reductase which interacts with the 2-amino pyrimidone ring of both molydopterin guanine dinucleotide molecules (Marchler-Bauer and Bryant, 2004).

The ArrA's from the three species for which the complete sequence is known (*Shewanella* sp. ANA-3, *D. hafniense*, and *W. succinogenes*) also have a leader sequence with a twin arginine motif, indicating a transported protein. Initial investigations of the arsenate reductase from the haloalkaliphilic Gram-positive bacterium *Bacillus selenitireducens* showed similar characteristics to ArrA from *C. arsenatis*. The ArrA from MLS10 is a 110 kDa protein, and shares 50% identity with *C. arsenatis*. It has an iron-sulfur cluster binding motif (C-X₂-C-X₃-C-X₂₇-C) in a conserved 36 aa sequence **C-Q-G-C-T-X-W-C-X₁₂-K-X₁₄-C** at the N –terminus, and a conserved 15 aa sequence at the C-terminus **K-C-Y-G-Q-G-H-X-A-X-G-X₃** and an internal molybdenum binding domain. A comparison of the amino acid sequence of ArrA from *B. selenitireducens*, *Shewanella* sp. ANA-3 and *W. succinogenes* shows that they share a 34% identity and a

64% similarity, which increases in pair wise comparisons, suggesting that the enzyme is conserved across diverse phylogenetic species.

2. *ArrB –the Fe-S cluster containing subunit*

ArrB is the smaller subunit of the arsenate reductase enzyme and contains a typical cysteine rich motif for example -(C-X₂-C-X₂-C-X₃-C), and four such clusters indicate a 4Fe-4S cluster binding motif which is conserved. The arrangements of the cysteines in these clusters differ between species and accordingly their predicted structures also vary.

Amino acid sequence comparison of ArrB from *Shewanella* sp. ANA-3, *D.hafniense*, and *W. succinogenes* and the first 30 amino acids from *B.selenitireducens* shows that they share a 43% identity and a 68% similarity. Also, a recognizable leader sequence in the N-terminus is absent in these sequences. From genetic analysis it is suggested that ArrB is transcribed, assembled, and transported along with ArrA (Berks, 1996; Saltikov and Newman, 2003).

3. *Gene order*

The arsenate reductase genes are encoded by the ‘*arr*’ operon. There is operon heterogeneity, where *Shewanella* ANA-3 has been reported to have only two open reading frames - *arrAB* operon (Saltikov and Newman, 2003) and *D. hafniense* has at least seven genes where *arrSKRC* are located upstream of *arrA* and *arrD* is located downstream of *arrB* (genome not yet complete). *W. succinogenes*, has an *arrC* upstream of *arrA*. *B.selenitireducens* has been sequenced for the *arrA* (Afkar et al., 2003) and

arrB genes (This study) (Figure 5).

4. *Bacillus selenitireducens* – Physiology

Bacillus selenitireducens strain MLS10 was one of two anaerobic haloalkaliphiles isolated from Mono Lake, an alkaline (pH 9.8), moderately hypersaline (salinity 90 g l⁻¹), arsenic rich (~300uM) water body; a soda lake located in an arid basin in eastern California (Switzer Blum et al., 1998).

MLS10 is halotolerant and surprisingly unlike *Bacillus arseniciselenatis* strain E1H, the other isolate from Mono Lake, can grow at salinities in excess of 200 g l⁻¹ which is more than twice the salinity of Mono lake (Switzer Blum et al., 1998).

B.selenitireducens strain MLS10 is a very valuable candidate for bioremediation of heavy metal/metalloid [Selenate (Se VI), Selenite (Se IV), Arsenate (AsV) and Arsenite (AsIII)] contaminated brines, aquifers and other wastewater bodies, the reasons for which are as follows. MLS10 grows by dissimilatory reduction of Se (IV) to Se (0) and As(V) to As(III) coupled to the oxidation of lactate to acetate + CO₂ . Energy is derived through this process. Pyruvate is the other electron donor. It has the capacity to use a wide range of electron acceptors such as fumarate, nitrate, nitrite and trimethylamine oxide in addition to the two substrates mentioned above for respiration. A weak microaerophile, MLS10 can tolerate and grow under low levels of O₂ as well as carry out fermentative growth utilizing glucose as the carbon source. *B. selenitireducens* grows at a range of pH 8 -11 (optimally at pH 9.8), can tolerate high salinity >200 g l⁻¹ and concentrations of arsenic e.g. 10mM (Mono lake arsenic concentration ~300uM) (Afkar et al., 2003).

5. Morphology of MLS10

B.selenitireducens is a short (2.0 – 6.0 μm in length, 0.5 μm in width) Gram positive, non spore forming rod. When grown with Se(IV) as the terminal electron acceptor, it forms small spheres of selenium on the periphery of its cell envelope. Later, this is expelled out into the medium of growth and dark red deposits of selenium or black deposits of elemental selenium can be observed which is sometimes internalized as well (Oremland et al., 2004).

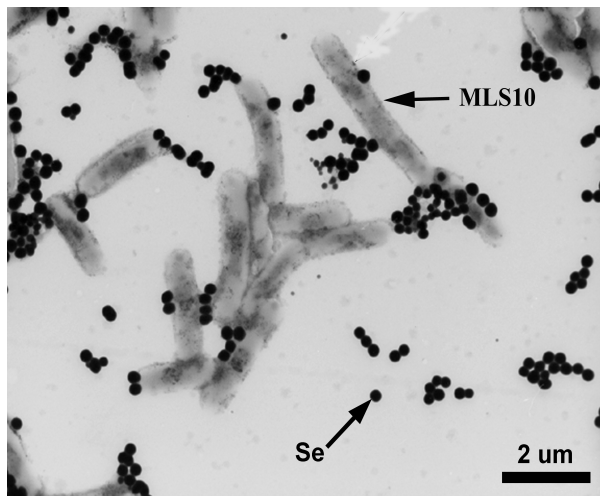


Figure 6: Whole mount (TEM) of a cluster of cells of *B.selenitireducens* strain MLS10 (dark spots indicate elemental selenium) (Oremland et al., 2004)

II. RESEARCH GOALS

Earlier, part of the *arrA* gene from MLS10 had been cloned and 2424 bases were sequenced with the 5' end incomplete and lacking ~180 bases (Afkar et al., 2003). A further investigation was also undertaken to sequence the downstream region of *arrA* and to look for a probable upstream *arrC* gene. This study was focused on the β (ArrB) subunit of the arsenate reductase from MLS10, its amplification, cloning and sequencing. To delve into this effort was very crucial, because, any information on the *arrB* gene sequence would throw light on better understanding of the arsenate reductase; generate sequence data for mutagenesis and thereby further the present understanding of the pathways involved in respiration aided by metalloids like selenate and arsenate.

This study also probed into a fairly recently isolated *Clostridium* sp. strain OhILAs, from the Ohio River enrichments. The specific aim of this project was to characterize the physiological growth conditions of OhILAs, as well as, perform enrichments of Ohio River sediments to examine other arsenate reducing bacteria (Edward Fisher, Masters Thesis). Professor Mahmoud Berekaa, had initiated sequencing from OhILAs genomic DNA in Spring 2005. My second research goal was to sequence the arsenate reductase genes from *clostridium* sp. OhILAs namely the *arrA* and *arrB* genes.

Hypothesis 1:

The *arrB* gene in MLS10 is a small Fe-Sulfur cluster subunit.

Specific aim

To clone and sequence the *arrB* gene from *B. selenitireducens* strain MLS10

Hypothesis 2:

Clostridium sp. strain OhILAs is a fresh water Gram-positive bacterium and has an arsenate reductase Arr similar to *Desulfitobacterium hafniense*.

Specific aim

To sequence the arsenate respiratory reductase '*arr*' genes from *Clostridium* sp. OhILAs, namely *arrA* and *arrB*.

III. METHODS AND MATERIALS

Cell culture and DNA extraction are described first followed by primer design, PCR optimization and amplification, cloning and transformation and finally, sequencing. The solutions and media used for the experiments and all the protocols are described in the end, under appendix.

Cell culture and DNA extraction

Bacillus selenitireducens strain MLS10 (ATCC 700615) was cultivated at 30°C in a mineral salt medium as previously described (Switzer Blum *et al.*, 1998) with sodium arsenate (10 mM) as the terminal electron acceptor and sodium lactate (15 mM) as the electron donor and (0.5%) yeast extract (Appendix 1). The pH of the media was raised to 9.8 and the medium transferred into Wheaton bottles (125 ml) containing 100ml of media. Each bottle was degassed with 80:20 N₂:CO₂ for 5 minutes for the liquid and 2 minutes for the headspace to ensure anaerobic conditions. Bottles were sealed with butyl rubber stoppers and capped with aluminum crimp tops before autoclaving.

Clostridium sp. OhILAs was grown at 37°C in a minimum mineral medium (Appendix 2) containing 20mM sodium arsenate as electron acceptor and 20mM lactate/acetate as electron donor and 1g yeast/L. The pH of the media was raised to 7.5 and the medium transferred into Wheaton bottles (125 ml) containing 100ml of media. Each bottle was degassed with 80:20 N₂:CO₂ for 5 minutes for the liquid and 2 minutes

for the headspace to ensure anaerobic conditions. Bottles were sealed with butyl rubber stoppers and capped with aluminum crimp tops before autoclaving.

The medium, at room temperature was inoculated with 2-3ml cells from an older MLS10/OhILAs culture as the case may be. Cells were harvested at an early exponential phase (after 48 hrs) by spinning down 100mls of cells divided into two-50ml centrifuge tubes at a speed of 7500 x g for 15 minutes. The supernatant was discarded and the cell pellet washed in 25ml of lysis buffer and spun down again at 7500 x g for 15 minutes. The supernatant was again discarded and the cells resuspended in 8ml lysis buffer. 900ul lysozyme was added and the reaction incubated at 37°C for 10 minutes without shaking. The reaction was removed from the 37°C incubator followed by the addition of 600ul of 20% SDS and mixed well. To this, 8ml Tris-buffered phenol was added and the resulting reaction mixed by inverting 4 times followed by centrifugation at 12,000 x g for 5 minutes. This forced the solution to separate into an upper aqueous phase containing the DNA and a lower phenolic phase. The upper aqueous was removed very carefully into a clean sterile 15ml centrifuge tube making sure that no portion of the phenolic phase was pipetted. DNA was precipitated by addition of 400ul of 5M NaCl and 20ml cold 100% ethanol. The reaction was left at -20°C for about 30 minutes and then the sample was spun down at 12,000 x g at 4°C for 10 minutes to retrieve the DNA. The resulting DNA pellet was washed with 70% cold ethanol and the pellet dried in a speed vacuum at medium heat (60°C). DNA when completely lyophilized is stable at room temperature. When resuspended in 200ul sterile water or TE buffer, DNA was always stored at -20°C (Appendix 3).

The extracted genomic DNA was treated with RNase A (10mg/ml in 10mM Tris pH 7.6, 15mM NaCl) according to RNase treatment of Genomic DNA protocol (Appendix 4). The genomic DNA was run on a 0.8% low melting agarose gel to check for a single band of DNA ~ 2-4Mb. All DNA was resuspended in ddH₂O and stored at -20°C after quantifying and estimating the purity, with the spectrophotometer (260/280 nm, 50ug DNA = O.D. of 1).

PCR and amplification with existing primers

The quality of the genomic DNA extracted from MLS10 was checked by amplifying for *arrA* using forward and reverse degenerate primers designed earlier for MLS10 *arrA* (Table 1, # 1 & 2, Figures 7 & 8) (Afkar et al., 2003).

Gene fragments were generated by PCR, using RedMix Plus (PGC scientific corp., Gaithersburg, MD,USA). Thin walled 0.2ml PCR tubes were used for all PCR reactions with the Techgene thermocycler (Techne incorporated, Princeton, NJ, USA).

A reaction mixture assembled on ice contained 50ng of *B. selenitireducens* DNA, 25pmol each of the forward and reverse primer, 12.5ul of RedMix Plus and ddH₂O to a total volume of 25ul.

The amplification parameters consisted of a 2 minute initial denaturation step at 94°C, followed by 30 cycles of denaturation for 30 seconds, annealing for 30 seconds at 50°C and an extension for 3 minutes (1min/kb of the gene or gene fragment) at 72°C. A final, extension step at 72°C for 10 minutes, was carried out for all samples that were to be used in cloning as the next step.

When amplified with Pfu Turbo polymerase (Stratagene, La Jolla, CA, USA), the final extension step was carried out by incubating the reaction product at 72°C for 10 minutes after the addition of 1 U of Taq DNA polymerase (Promega, Madison, WI, USA).

This was done to aid the addition of Adenine overhangs, since, Pfu Turbo polymerase produces blunt end PCR products, which will not facilitate ligation into the vector during cloning and transformation. All products obtained, were analyzed by 0.8% agarose gel in 1X TAE buffer.

Amplification of the Fe-S cluster subunit - *arrB* from MLS10

The two forward primers (34JS F3 and 34JS F4) had been designed based on the 3' nucleotide sequence of *arrA* (GenBank accession number AY283639) (Afkar et al., 2003) and another 340 bases immediately downstream (Lisak and Stolz, Unpublished results) respectively (Table 1, #3&4, Figure 8).

The reverse primers were sequence specific, designed using the Primer 3 program (Lisak and Stolz, Unpublished results) based on the nucleotide sequence of *arrB* 3' end from *Shewanella* sp. ANA-3 (ArrBSh R1 and ArrBSh R2) and *Desulfitobacterium hafniense* (ArrBDh R3 and ArrBDh R4). These, 2 forward and 4 reverse primers (Table 1, #5-8, Figure 7) were used for amplification and subsequent sequencing of the *arrB* gene from MLS10, utilizing Pfu Turbo polymerase, which is recommended for amplification of gene fragments bigger than 1kb.

The reaction mixture assembled on ice contained 50 ng of *B. selenitireducens* genomic DNA, 1ul each of forward and reverse primers from a 25pmol stock, 0.5ul of

Pfu Turbo polymerase, 2.5ul of the 1X reaction buffer, 0.5ul of dNTPs from a 200uM stock and ddH₂O to a final volume of 25ul.

The amplification parameters consisted of a 5 min denaturation step at 94°C, followed by 30 cycles of 94°C, 30 seconds; 50°C, 30seconds; and 72°C, 3 minutes. A 3' A- extension was added by incubating the reaction product for 10 minutes at 72°C after the addition of 1 U of Taq DNA polymerase for reasons mentioned above.

Purification of PCR product

PCR products were purified to remove any unincorporated nucleotides, primers and small DNA fragments. Samples were separated on a 0.8% low melting agarose gel for 90 minutes at 100 volts to allow for clear isolation of fragments. Bands of the expected size were excised with a clean fresh razor blade and the DNA eluted using the GenElute Agarose Spin Columns (Sigma Aldrich, St. Louis, MO, USA).

To proceed with direct sequencing as per the protocol (Appendix 5), it was critical to have high concentrations of pure PCR products. This necessitated setting up several identical PCR reactions which would facilitate easy elution from the agarose gel and the resulting product would be highly concentrated. The products were cleaned with new CENTRI-CEP spin columns (Princeton Separations, Adelphia, NJ, USA) or used columns refilled with 800ul Sephadex G-50 slurry (Appendix 5). This ensured the removal of impurities in the form of unincorporated DNA, primers or smaller fragments of amplified DNA and the DNA was checked for product recovery by running on a 0.8% agarose gel under identical conditions mentioned earlier.

The concentration of DNA was estimated by spectrophotometer and 40 – 100ng of the DNA was used for preparing the sample for direct sequencing with the BigDye Reagent (Applied Biosystems, Foster City, CA,USA) using the forward/reverse primers, as the case may be (Appendix 5).

Primer design

Primers were designed based on the following guidelines:

1. Length of the primers was between 17-22 nucleotides such that the T_m was manually calculated $[2(A/T) + 4(G/C)]$ to lie between 50°C - 58°C. Every degenerate base was also given a score of 2, for T_m calculation.
2. Forward primers were based on the plus strand and reverse primers on the minus strand .
3. Any possibility for “Primer dimer” (dimerization with itself), or the forward/reverse primer was minimized.
4. GC rich 3' ends were avoided.
5. Degenerate primers were designed based on sequence alignments from several organisms related to gene of interest.
6. Primers were purified by “standard desalting” and ordered from Integrated DNA technology Inc., Coralville, IOWA, USA.

New primers for amplification of MLS10 arrB gene

New primers were designed for amplification of *arrB* from MLS10. The strategy was to use forward primers based on the 3' end of *arrA* (Figure 8), and new degenerate primers based on conserved sequences in *arrB* (Figure 7).

The primers were designed manually following the guidelines outlined under primer design. Three forward primers – MArrB F1, MArrB F2 and MArrB F3 (MLS10 *arrA* sequence specific) were designed such that they were positioned at the 3' end of MLS10 *arrA* (Table 1, #9-11, Figure 8). The 4 degenerate reverse primers – DArrB R1, DArrB R2, DArrB R3 and DArrB R4, were designed based on *arrB* multiple sequence alignment (Clustal W 1.82) from *Desulfitobacterium hafniense*, *Wolinella succinogenes* and *Shewanella sp.* ANA-3 (Table 1, #12-15, Figure 7).

Primers used for the amplification of arrAB gene from Clostridium sp. OhILAs

Degenerate and species specific primers were designed based on multiple sequence alignment of *arrA* sequences from MLS10, *B.arseniciselenatis* strain E1H, *Sulfurospirillum barnesii* SES-3, *Shewanella sp.* strain ANA-3, *Desulfitobacterium hafniense* and *Wolinella succinogenes*, to test for use in Denaturation Gradient Gel Electrophoresis (DGGE) (Berekaa and Stolz, unpublished results). These included 19 primers (5 forward, 4 reverse, 5 degenerate forward and 5 degenerate reverse primers) (Table 1, #17-35, Figures 8 & 9). The MLS10 *arrA* sequence served as a reference template.

Following these, were the primers that had been designed for MLS10 *arrB* amplification. Thus, a combination of the two sets of primers was used to amplify the *arrA* and *arrB* gene from OhILAs genomic DNA.

PCR optimization and amplification

MLS10 *arrB* amplification primers (Table 1) had a range of melting temperatures (47.6 to 62.0°C). The size of expected fragment also varied, depending on the combinations of the forward and reverse primers (375bp to 1200bp). This high variability required diligent and comprehensive reading of results from several PCR reactions with varying annealing temperature to suite every special primer pair and an extension period (72°C) to suit the size of the expected fragment. PCR optimization, therefore, played a major role in amplification and sequencing of expected fragments of DNA from the starting material i.e. MLS10/OhILAs genomic DNA.

The case of OhILAs was no different, with the goal focused on amplifying and sequencing two successive genes *arrA* and *arrB* from the respiratory arsenate reductase *arr* operon.

The cycling parameters were one of three templates and they were as follows
The cycling parameters used for this reaction was as follows:

- | | | | |
|----|-----------------------------|--------------------------------|-----------|
| A. | 1. Initial denaturation | 95°C | 5 minutes |
| | 2. Denaturation | 95°C | 1 minute |
| | 3. Annealing | Varying Temp; primer dependant | 1 minute |
| | 4. Extension | 72°C | 1 min/kb |
| | 5. steps 2 to 4 – 35 cycles | | |

B.	1. Initial denaturation	94°C	5 minutes
	2. Denaturation	94°C	1 minute
	3. Annealing	Varying Temp; primer dependant	1 minute
	4. Extension	72°C	1 min/kb
	5. steps 2 to 4 – 35 cycles		

C. Touch Down PCR

1. Initial denaturation	94°C	10 minutes
2. Denaturation	94°C	1 minute
3. First temp.	52.5°C	0.5°C fall/cycle
4. Last Temp.	42.5°C	1 minute
4. Extension	72°C	1 min/kb
5. Step 2 – 15 cycles		
6. Step 3 – 25 cycles		

Cloning of the PCR product

The PCR fragment of interest, that contained the largest part of the gene to be sequenced, was amplified with Pfu Turbo in several identical reactions followed by gel

separation and purification as mentioned earlier. The 3' A-extension was then added by incubating the reaction product for 10 min at 72°C after the addition of 1 U of Taq DNA polymerase just before beginning the protocol for cloning. The pCR -4- TOPO kit was used for cloning and transforming the PCR product into chemically competent *E.coli* cells, according to the manufacturer's instructions (Invitrogen Carlsbad, CA, USA).

After growing transformants at 37°C, overnight on LB 100µg ml⁻¹ ampicillin /50µg ml⁻¹ kanamycin agar plates (Appendix 6) (depending on the size of the insert), selected colonies were picked for further analysis and cultured overnight in 1-10ml liquid media (Appendix 6) containing the same antibiotic as the agar plates. Sterile conditions were maintained throughout the proceedings of the protocol to avoid cross-contamination in the plates and liquid media.

Plasmid purification and analyzing transformants

The 1-10ml overnight cultures were purified by utilizing Wizard *Plus* SV Minipreps DNA Purification System (Promega, Madison, WI, USA), recommended by Applied Biosystems. Digestion with the restriction endonuclease EcoR1 was done next, to analyze for the presence of insert (PCR product) in the vector.

The reaction assembled on ice, contained 1ul of EcoR1 restriction endonuclease, 2ul of the 10X buffer H, 10ul of the plasmid DNA and 7ul of ddH₂O to a total volume of 20ul.

The reaction was gently mixed and incubated at 37°C for 2 hours. The reaction was then stopped by adding 0.4µl of 0.5M EDTA, pH 8.0 or alternatively by freezing the reactions at -20°C. The digests were then separated on a 0.8% low melting agarose gel to analyze for presence of inserts.

EcoR1 cuts the circular (vector + insert) DNA at two specified points on the vector sequence, flanking either side of the PCR insert, and could make none or several cuts on the insert DNA. Depending on the number of restriction sites, the digest, when run along with the undigested plasmid DNA on a 0.8% agarose gel gives clear suggestions on whether the PCR product of interest indeed was successfully inserted into the vector. Another method to check for the presence of inserts was to do 'The Cracking Procedure' (Appendix 7), which basically facilitated breaking open the vector, such that the insert and the vector DNA are linearised into two fragments and the reaction was then run on a 0.8% agarose gel to look for the insert and the vector (3957 bp; Invitrogen manufacturer's manual), whose sizes are known.

Sequencing from PCR product / plasmid DNA

The concentration of DNA was estimated by spectrophotometer and 40 – 100ng of the DNA was used for preparing the sample for direct sequencing with the BigDye Reagent (Applied Biosystems, Foster City, CA, USA) using the forward/reverse primers (Appendix 5).

Protocol for Direct Sequencing of PCR products:

PCR product >2000bp	40-100ng
Big Dye Reagent	2 μ l
2.5X Sequencing Buffer	6 μ l
Primer	1 μ l (from 20uM stock)
ddH ₂ O	to a total volume of 20 μ l

The cycling parameters for this reaction were as follows:

1. Initial denaturation 95°C 1 minute
2. Denaturation 95°C 10 seconds
3. Annealing 50°C 5 seconds
4. Extension 60°C 4 minutes
5. steps 2 to 4 – 55 cycles

The sample was again purified with the CENTRI-SEP columns, lyophilized and resuspended in 10 μ l DI formamide according to the protocol (Appendix 5) before sequencing. Sequencing was carried out on an ABI 3100 sequencer at the sequencing facility in Duquesne University. Each sequencing reaction gave 550-750bp of readable sequence.

Pure plasmid DNA containing the insert was quantified by spectrophotometer and appropriate quantities used as template (Appendix 5) with M13 primers (pCR2.1 TOPO

kit) or the PCR primers to sequence the gene of interest. Also, the successful clone was plated and stored at 2-8°C for future plasmid preparations and plasmid DNA extraction.

20ul of the sequencing reaction contained 300 ng of Plasmid DNA (eluted in water), 2ul of the Big Dye reagent, 6ul of the 2.5X Sequencing buffer, 1ul of the Primer (forward/reverse) from a 20uM stock and appropriate quantity of ddH₂O to a total volume of 20ul.

The cycling parameters for this reaction were as follows:

1. Initial denaturation 95°C 1 minute
2. Denaturation 95°C 10 seconds
3. Annealing 50°C 5 seconds
4. Extension 60°C 4 minutes
5. Steps 2 to 4 – 55 cycles

As mentioned earlier the reaction was purified according to the protocol (Appendix 5) and then sequenced. Each sequencing reaction gave 550 – 750bp of readable sequence.

Sequence analysis

The sequence data was used for predicting Open Reading Frames (ORFs) with the Orf finder program in GenBank. The sequence was also used for BLAST searches (www.ncbi.nlm.nih.gov –‘blastn’, ‘blastp’, ‘blastx’) to look for close matches of genes/proteins in the database. blastn gave close nucleotide matches, blastp generated closest protein matches and blastx provided the closest protein matches for the nucleotide query sequence. The predicted molecular mass and pI are also calculated using the program calculate MW/pI at the website (www.expasy.org). The program SIGNALP (www.cbs.dtu.dk/services/SignalP/) was used to identify a Tat recognition region, which is suggestive of the protein being secreted. ClustalW 1.82 and ClustalX 1.83 were used for creating all alignment files for both the nucleotide and protein sequences.

	PRIMER	SEQUENCE	REFERENCE
1.	110JS F1	GAR CAR GGN GAR TGG ATH GC T _m = 57.5 °C, Length = 20	(Afkar <i>et al.</i> , 2003)
2.	34JS R2	DAT NGT CAT NGC RTA RTT YTT YTT T _m = 53.1°C, Length = 24	(Afkar <i>et al.</i> , 2003)
3.	34JS F3	CGA AAA GGT GAA TTG GCA TT T _m = 53.1°C, Length = 20	(Afkar <i>et al.</i> , 2003)
4.	34JS F4 *	TAT CCG CTC ACA ATT CCA CA T _m = 55.2°C, Length = 20	(Afkar <i>et al.</i> , 2003)
5.	ArrBSh R1	ACC AAT TAC CCG TGC TTC AG T _m = 56.3°C, Length = 20	Lisak and Stolz unpublished results
6.	ArrBSh R2	AAA TAC CCT TGG TCC CGT TC T _m = 55.8°C, Length = 20	Lisak and Stolz unpublished results

7.	ArrBDh R3	TGA CTG ACT TCG CTT GTT GG T _m = 56.4°C, Length = 20	Lisak and Stolz unpublished results
8.	ArrBDh R4	GGC TGA GTT CCA GCA TTA GG T _m = 56.4°C, Length = 20	Lisak and Stolz unpublished results
9.	MArrB F1	TTT GAG TTT TAC AGT GAG AC T _m = 47.6°C, Length = 20	This study
10.	MArrB F2	GCA AAA GGT GAA TTG GCA TT T _m = 52.3°C, Length = 20	This study
11.	MArrB F3	ATG TGG CAA AAT TGA ATC CA T _m = 50.7°C, Length = 20	This study
12.	DArrB R1	AGS CKC CAC ABC CHA CRC AT T _m = 62.0°C, Length = 20	This study
13.	DArrB R2	ACA CAD GSW GCA TYA TCR CA T _m = 55.6°C, Length = 20	This study
14.	DArrB R3	RCA VYT CTT RCA VCC RAT GCA T _m = 57.5°C, Length = 21	This study
15.	DArrB R4	GGA TCA TYB ARA TCV CCA A T _m = 50.4°C, Length = 19	This study
16.	MarrX F1	ATG CAG AAT TAG GAG AGT TC T _m = 49.4°C, Length = 20	This study
17.	arrAMLSF1	TGT CAA GGT TGT ACG GCA TG T _m = 55.4°C, Length = 20	Berekaa and Stolz Unpublished results
18.	arrAMLSF3	ACG CCT GAA TGG GCA GCT T _m = 60.3°C, Length = 18	Berekaa and Stolz Unpublished results
19.	arrAMLSF4	AGA CCC TGA AAC GGG TGA A T _m = 56.2°C, Length = 19	Berekaa and Stolz Unpublished results
20.	arrAMLSF5	AGT TTG AGT TTT ACA GTG AGA C T _m = 50.5°C, Length = 22	Berekaa and Stolz Unpublished results
21.	arrAUF1	TGT CAA GGH TGT ACB DCH TGG T _m = 56.2°C, Length = 21	Berekaa and Stolz Unpublished results
22.	arrAUF2	TGG GAN GAA GCD ATB GAT AC T _m = 53.1°C, Length = 20	Berekaa and Stolz Unpublished results
23.	arrAUF3	CAN CCB GAR TGG GCW GC T _m = 60.3°C, Length = 17	Berekaa and Stolz Unpublished results

24.	arrAUF4	ARA GAY CCH GAA ACB ACA TC T _m = 52.0 °C, Length = 20	Berekaa and Stolz Unpublished results
25.	arrAUF5	ADT TTG AGT TYT AYA GYG AAA C T _m = 48.7°C, Length = 22	Berekaa and Stolz Unpublished results
26.	arrAUF6	GCC TTT GTN CCN CAC TAT GA T _m = 54.9 °C, Length = 20	Berekaa and Stolz Unpublished results
27.	arrAMLSR2	GTA TCC ATT GCT TCT TCC CA T _m = 52.3°C, Length = 20	Berekaa and Stolz Unpublished results
28.	arrAMLSR3	AGC TGC CCA TTC AGG CGT T _m = 60.3°C, Length = 18	Berekaa and Stolz Unpublished results
29.	arrAMLSR4	TTC ACC CGT TTC AGG GTC T T _m = 56.2°C, Length = 19	Berekaa and Stolz Unpublished results
30.	arrAMLSR8	GGA TTC AAT TTT GCC ACA TCA T T _m = 52.1°C, Length = 22	Berekaa and Stolz Unpublished results
31.	arrAUR2	GTA TCV ATH GCT TCN TCC CA T _m = 53.1°C, Length = 20	Berekaa and Stolz Unpublished results
32.	arrAUR3	GCW GCC CAY TCV GGN GT T _m = 60.3°C, Length = 17	Berekaa and Stolz Unpublished results
33.	arrAUR5	GTT TCR CTR TAR AAC TCA AAH T T _m = 48.7°C, Length = 22	Berekaa and Stolz Unpublished results
34.	arrAUR6	TCA TAG TGN GGN ACA AAG GC T _m = 54.9°C, Length = 20	Berekaa and Stolz Unpublished results
35.	arrAUR8	GGG TTV AWT TTN GCB ACA TC T _m = 51.5°C, Length = 20	Berekaa and Stolz Unpublished results
Where R = A,g; Y = C,T; M = A,C; K = g,T; S = g,C; W = A,T; H = A,C,T; B = g,T,C; V = g,C,A; D = g,A,T; N = A,C,g,T			

Table 1: Primers used in this study. * 34JSF4 was wrongly designed based on vector sequence.

1 **GAACAGGGCGAGTGGATCGC**AAGCGTT**TGTCAAGGTTGTACGGCATG**GTGTGCTGTTCA 110JSF1, [arrAMLSF1](#)
 AGTCTACAGAATTGATGGACGAGCAACAAAGGTCAGAGGTAATCCGAATGCGAAAGCGA
 ATCATGGTCATTCTGTGTGCGTTCACATATTGGACTGCAGCAAGTATACGATCCGGAC
 CGCGTAAAGCAGCCGATGAAGCGTACAAATCCAAACAAAGGCCGCGATGAAGATCCGGA
 ATTTGTACCGATCAGCT**TGGGAAGAAGCAATGGATAC**GATTGCCGATAAAATCATTGAAC [arrAMLSR2](#)
 TTCGTGAGAATAACGAAACACACAAATTCAGCGTTTGGCGCGGACGTTATACATCTAAC
 AATGGCATTCTCTACGGAAATATGCCAAAGATTATCGGTTACCAAAATAATATTTCTCA
 CAGCTCCATCTGTGCTGAATCCGAGAAATTCGGACGCTATTACACAGAACGCTACTGGG
 GTTATGCTGACTACGATCATGAAAATGCTCTTTATGAAATCTTCTGGGGCGGCGATCCG
 ATTGCAACAAACCGTCANGTGCCACATACGGCGAGTATTTGGGGAGAGTTGTCCGACAG
 AGCAACGCTTGCATGTGTAGACCCACGATTCTCAACAACGCTGCTAATCAGATGAATG
 GATGCCGGCTTATCCGGGGGAAGACGGTGCGATTGCATCTGCTATTGCGCACGTTATC
 TTAACCTGAAGCGCTTTGGTATAAGCCGTTTCGTGCGTGACTTCAAAGATGGACAAAACCG
 ATTCGTTGAAGGCCGCGATGTGAATGAAGATGACTTTGAAGAAATTCAGACACACGGAC
 TTGTAAATGGTGAATCTAGAACTGAAAGATAAG**ACGCCTGAATGGGCAGCT**GAACGC [arrAMLSF3/R3](#)
 TCCGGTATCGATGCTGATCAAATTTACCGCGTAGCAAGAGGATTGCAAATGCTGCACC
 AAAGGCAATCTCGTTCTCATCACCAGGATCAAGTATGACAATCCGCGGCGGTTACACAG
 CGATGACACAAGCAGCCTTAAACGGCCTTGTCGGTTCTGCAGATAACTTAGGTGGAGTC
 ATCAGTAATGGGATCTCTGTTCCAAATAACGGCTTCCCGGATCCATCCGATTATATTGA
 TGAGATTGCAGCGAATGGTTTAGAGCAAGACCGCATTGACTGGGGAGGAAGGCTGGAAT
 TCCCTGCACTGAAAGATAGCAAATCCGGTGGGGTTAAAGTAACGAATACCGTTGCTGAT
 GCCGTACTTGCAGAAGATCCTTACGAATTTAAAAGTCGTATTGAGCTATTGGACGAACT
 TCAATTTCTCCAACCAAGGAACGATCGTTGGGACAGAGCTTGGCCAAGATTCCATTTC
 ATGGTGCATATGACTGTCAATCCGGCGAACAACCGCATTTTGCAGATATTGTCCTTCCG
 GTTCCGCACTCGCAGTTTGAGCGTCAAAGCCCGGTTGCAGGCAGTAACGGTAATCTCCA
 CAGACACTTGCATTTGCAAAATAAAGTCATTGAATCACCATTGATATCCGGGTTGATG
 AAAGTGAATTTCCATGGATGATCGGTGAATCACTTGAGAAAAAAGGATATTCCAACCTG
 ATTGATTATTTCCGTAATGAATTCAG**AGACCCTGAAACGGGTGAA**GCACCAACAAACGC [arrAMLSF4/R4](#)
 CGCAGAATTTCGATGAAATTGCAACGAAACATTACACACATCCAGTCTGGGACCCAACCT
 TTGAAAAAGATGGCGATCAGATCGATGGTTGGGAAGAATATAAGCGATTGGGAACATGG
 AACACCAATAAATATGAGTTCCGTCAAAAGTGGGATGGAAATTGGGGAACAGAAACCGG
 AC**AGTTTGAGTTTACAGTGAGAC**TTTAAAGTGGCATTACAAGATCATGCAGACAAAC [arrAMLSF5, MArrBF1](#)
 ACAATGCAAGCATTGATGATGTTATGGAGGCCACATTCAATACA**GCAAAAGGTGAATTG** 34JSF3/MArrBF2
GCATTTGTACCTCACTATGAACCGGCGATGCGTGTAGGTGACGAATCCGAATATCCGCT
 GATCTTCATGGAACACCGTTCAAACTGAATCGTGAAGCGCGATCAGCAAATACAAGTT
 GGTACCAGGAATTTAAGGATATCGACCTTGGTGATGAAGCATGGGAT**ATGTGGCAAAA** [MArrBF3, arrAMLSR8](#)
TTGAATCCAAAGGATGCTGCAGAACTTGGTATTCAAAACGGTGATATGGTCCGTTGGT
 TACGCCTGAAGGTCAAATCGAAGTCAAAGCCAAGCTCTGGGAAGGTACACGTCCAGGAG
 TTGTTGCCAAATGTTACGGACAAGGTCCTGGGCATATGGCCATATCGCTTCACTTGAT
 CGTCGCCGTCAAATCGCCAGAGGCGGGAACAACAACATTATCCTCGCTCCTGTGCACGA
 AGCATTAAAGCGGAAGCGGTGCAAGACACGGTGGACAGACCCGAGTAAGAGTAGAGAAAG
 TATAAGTTAGAGACATTATGCGGAGGTGACACAAATGGCT**AAGAAAAATTATGCAATGA** 34JSR2
CGATAGATTTACAGGCGTGATCGGCTGTGCCGCTGTGCGGTCACTTGTA AAAACGAG
 AAC 2540

Figure 8: Position of primers 1-3, 9-11, 17-20 and 27-30. *Bacillus selenitireducens* strain MLS10 respiratory arsenate reductase catalytic subunit (*arrA*) gene – 2424 bases (AY283639.1) followed by an intergenic 29 bases and 87 bases of *arrB* gene 5' end.

CLUSTAL W (1.82) multiple sequence alignment: *arrA* - most significant alignments

SES	G---GTGAATTACAGTACGAAGGTGAAA-----CTGGCAAATGGATGAGCTCCACC	150
Wolin	-----GCAGGTTTCAG-----GCGAAG-----TGGGGAAATGGGTCGCTAGCACC	132
MLS10	-----GAAC-----AGGGCGAGTGGATCGCAAGCGTT	27
ElH	-----GAATTTAGCGGCC-----GCGAAT-----TCGCCCTTTGGATGGCCACCATG	67
Dhaf	-----GAAAAATACCGCCCTA--GTGCAG-----AAGGAAAATGGATGAGCACGCTC	147
ANA3	GCCGCTGAATTACCCGCGCCTCTACGTAGAACGGGCGTGGGTGAATGGCTAGCCACGACT	180
	* * * * *	
SES	TGTCAAGGCTGTACCAGTTGGTGTCCGATTCAAGGTCTTGTGGTTGATGGAAAAGTCGTT	210
Wolin	TGTCAAGGATGCACTAGTTGGTGCTCTATTTCAGGGCTATGTTGTGGATGGGAGGCTGGTT	192
MLS10	TGTCAAGGTTGTACGGCATGGTGTGCTGTTCAAGTCTACAGAATTGATGGACGAGCAACA	87
ElH	TGTCAGGGCTGTACTGCTTGGTGTGCCGTTCAAGTTTATATAATTAATGGTCGCGCAACA	127
Dhaf	TGTCAAGGATGTACCCTGGTGTGCCATTCAAGCCTATGTCGTGGATGGCCGAGTGATT	207
ANA3	TGCCAAGGTTGTACCTCATGGTGCAGTAAACAGATTATGTTATGGATGGCCGCGCATTA	240
	** * * * * * * * * * *	
	TGTCAAGGHTGTACBDCHTGG	arrAUF1
SES	AAAAGGCAAAGGCATTGATCCTAAATTCGTGCCCATTTCTGTGGGACGAAGCGATTAATAC	392
Wolin	AAAAGGAAAAGGAATCGACCCTCAGTTTGTCCCCATTAGCTGGGATGAGGCAATCGATAC	377
MLS10	CAAAGGCCGCGATGAAGATCCGGAATTTGTACCGATCAGCTGGGAAGAAGCAATGGATAC	272
ElH	AAAGGGGAGAGATGAAGATCCAGGGTTTATACCAATTTCTTGGGAAGAAGCAATGGATAC	312
Dhaf	GAAAGGGCGTAAATGAAGATCCCAATTCGTCCCTATCACTTGGGAGGAAGCTATTGATAC	392
ANA3	AAAAGGCCGCGACCAAGATCCTAAATTTGTCCCTATTAGCTGGGATAAAGCGCTGGATAT	425
	* * * * * * * * * * *	
	TGGGANGAAGCDATBGATAC	arrAUF2
	ACCCCTNCTTCGHTAVCTATG	arrAUR2
SES	TGGTGGAACTTAGAATTAAGAGAGAACGCCTGAGTGGGCTGAAAAAATCACAGGAATT	1044
Wolin	TGGTGGAACTTTGAGCTCTTTGATAAGACACCCGAATGGGCTGAACCTATCACGGGGATT	1029
MLS10	TGGTGGAACTAGAACTGAAAGATAAGACGCCTGAATGGGCAGCTGAACGCTCCCGGTATC	894
ElH	TGGTGGAACTTGCATTGAAAGATGCAACGCCTGAGTGGGCAGCTGAAAGAAGTGGAAATT	934
Dhaf	TGGTGGAACTTAGAACTGAAGGATCGTACTCCGGAATGGGCTGCCGACGGGCAGGTATA	1014
ANA3	TGGTGGAAACCAGGCCCTTAAAGATTACCCCTGAATGGGCAAGCAAAATTACCGGTATT	1047
	***** * * * * *	
	CANCCBGARTGGGCWGC	arrAUF3
	TGNGGVCTYACCCGWCG	arrAUR3
SES	GATAACTTCAAAGACCCAGAAAGCGGTAAAGCACCCACAAACCCAGAAAGATTGACCTC	1809
Wolin	GAGAATTTCAAAGATCCTGAAACAGGAAGATTCCCTCCACGCCGAAAGAGTTCGATCTT	1794
MLS10	AATGAATTCAAGAGACCTGAAACGGGTGAAGCACCAACAAACGCGCAGAAATTCGATGAA	1669
ElH	AACGAGTTTAGAGATCCTGAAACTGGAAAACACCTGAAAATGGTCTTGAATTTAATGAG	1705
Dhaf	AACGAAGTCAAAGACCCGAAACCGGTAAAACTCCGAGCTCCACGAGGAATTCACCATG	1782
ANA3	GAACAAATGTTGATCCAGAAACCGGTAAACCTGCGGCGGATGAAGCTGAATTTGCTAAA	1818
	* * * * * * * * * * *	
	ARAGAYCCHGAAACBACATC	arrAUF4
SES	CAAAAAGGTTGAT---AACTTTGCAACCGAAACAAAAAAGTTTGAGTTTTACAGCGAAAC	1991
Wolin	CAAGAAGATTGAC---AATTTCAAACCGAAACCAAAAAATTTGAGTTTTATAGCGAAAC	1976
MLS10	TCAAAAGTGGGATGGAAATTGGGGAACAGAAACCGGACAGTTTGAGTTTTACAGTGAGAC	1854
ElH	TCAAAAGTGGAAATGTTAATTGGGGAAGTGAAGTGGCGATTTTGAGTTTTACAGTGAAC	1890
Dhaf	TAAAAACTGGGAT---AAATTCGGTACAAAAACCAAAATTTGAATCTATAGCGAAAC	1964
ANA3	AGCTCGTGGGGT---AAGTTTAAACAGAAACCACTAAATTTGAGTTCTACAGCAAAAC	2000
	* * * * * * * * * * *	
	ADTTTGAGTTYTAYAGYGAAC	arrAUF5
	THAACTCAARTRTCRCTTG	arrAUR5
SES	GGCATTAGAGGCGCTCAATTATACGTGTAGAGGAGAATTAGGCTTTGTCCCACTATGA	2105
Wolin	GGCGATTGAAGCTTCTAATAATACCGTAGGGGTGAGCGTGCTTTGTGCCGCACTATGA	2093
MLS10	TGTTATGGAGGCCACATTCAATACAGCAAAAGGTGAATTGGCATTGTACCTCACTATGA	1968
ElH	AGTAATGGAAGCGACGAATTACACTGCTCGTGAGAGAACTAGCGTTTGTTCACACTATGA	2004
Dhaf	TATTATGCAGGCAACCAATATTTAGCCAAGGGAGAACAGCCTTTGTTCCCTCACTACGA	2078
ANA3	GGTAATGAAAGCCTGTGACTACCAAGCCCGTGCCACTTAGCCCTTTATCCCCATTATGA	2114
	* * * * * * * * * * *	
	GCCTTTGTNCCNCACTATGA	arrAUF6
	CGGAAACANGNGTGATACT	arrAUR6

```

SES      GTTGACCCAGGTGATGTTGCAGGAAAAGATGTCACGAAAATCAACCCGCTGGATGGCAAA 2283
Wolin    GTGGATCCGGGTGATGAGATTGAGAAGGATGTGGCCAAGATTAATCCGAAACAGCTAAA 2271
MLS10    ATCGACCTTGGTGATGAAGCATGGGATGATGTGGCAAAATTGAATCCAAAGGATGCTGCA 2146
E1H      GCTGATCCTGGTGATGAGGCGTGGGATGATGTGGCAAAATTAATCCAATTGATGCAAAA 2182
Dhaf     GTAGACCCCGGTGATGAGGCTTGGGATGATGTAGCTAAAATCAACCCAGTGGATGGGGCA 2256
ANA3     GTCGATCCAGGTGATGTGGCCAATGAAGATGTGGCGAAATTCAACCCGATAGACGGTAAG 2292
          ** *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
          CTACABCGNTTTWAVTTGGG                                     arrAUR8

```

Figure 9: Position of primers 21-26 and 31-35. Clustal W(1.82) multiple sequence alignment of *arrA* gene sequences from *Sulfurospirillum barnesii* SES-3, *Wolinella succinogenes*, *Bacillus selenitireducens* strain MLS10, *B.arseniciselenatis* strain E1H, *Desulfitobacterium hafniense* and *Shewanella* sp. strain ANA-3.

IV RESULTS

PCR and amplification with existing primers: MLS10 arrA amplification

An expected fragment of ~2.5 kb was amplified from MLS10 genomic DNA with the primers 110JSF1 and 34JSR2 (Figure 27, Panel A). This could be achieved with the use of both Taq DNA polymerases (RedMix Plus) as well as with Pfu Turbo DNA Polymerase. Considering the fact that the expected gene (*arrA*) was bigger than 1kb, Pfu Turbo was recommended for amplification of genes greater than 1kb. After scrutinizing the PCR results on a 0.8% agarose gel and on comparing it with results obtained earlier, the DNA was eluted and purified for sequencing.

Direct sequencing vs. cloning, transformation and sequencing

Amplification of the catalytic subunit arrB from MLS10

The aim was to use, very recently designed *arrB* primers, which had not yet been tested extensively towards their purpose of design. Several combinations of the 6 primers (Table 1, #3-8, Figures 7 & 8) were tested with PCR reactions and two fragments were chosen for further investigation. They included a ~2.2kb fragment amplified with 34JSF4/ArrBShR2 and a ~2.5kb fragment amplified with 34JSF3/ArrBDhR3, both appearing as dominant bands (Figure 27). These fragments were followed up for sequencing.

The two dominant fragments mentioned above were prepared for direct sequencing and both the forward and reverse primers for the respective fragments i.e. either 34JSF3 (forward primer) or ArrBDhR3 (reverse primer) were used for the

sequencing reactions for the 2.5 kb fragment. After several trials of unproductive sequencing it was decided to clone both fragments into *E.coli*.

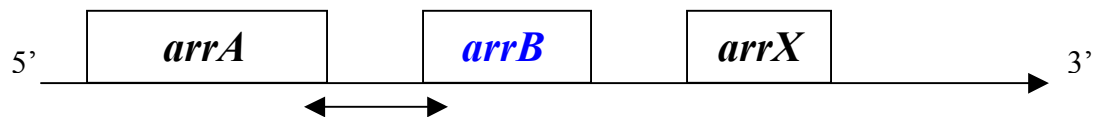
The bigger fragment of the two was cloned (Figure 27, Panel B) after the addition of A overhangs into pCR 4 TOPO (3957 bp) and transformed into chemically competent *E.coli* cells and the transformants were plated onto LB agar with 100ug/ml Ampicillin. After incubating overnight at 37°C the plates were examined and selected colonies were purified for further analysis. To perform the cracking procedure (Appendix 7), selected colonies were spread on the selective antibiotic plate, let to grow overnight at 37°C and used completely. This procedure clearly showed that we did not have our product of interest inserted into the vector. None of the clones that had been selected had the insert; clearly a ~2.5kb band, which was the size of the insert, was missing, and thus the first attempt at cloning MLS10 *arrB* into *E.coli*, had also failed (Figure 27, Panel C).

Direct sequencing from PCR product 34JSF3/ArrBDhR3 containing the arrB gene of interest

To proceed with direct sequencing as per the protocol (Appendix 5) it was critical to have good concentrations of pure PCR products.

The sequencer ABI 3100 was loaded with the lyophilized sequencing reaction resuspended in 10µl formamide. The quality of the results would directly reflect that of the sample provided. Thus, a very determined attempt was made to prepare all samples with utmost care, so as to negate any impurities that would hamper the production of clean readable good-sized sequences. Sequencing was attempted several times. The first 2 attempts with a total of 3 reactions gave around 600 –750 bases of readable sequence.

The first reaction with 80ng of PCR product (2.5 kb) as the template DNA and 34JSF3 as the sequencing primer gave 805 bases (Figure 10). BLAST search with these 805 bases gave the closest match as *Bacillus selenitireducens* strain MLS10 respiratory arsenate reductase catalytic subunit (*arrA*) gene bases 2001 –2424: bases 64 - 492 from the sequencing results and the second match as *Bacillus arseniciselenatis* arsenate respiratory reductase (*arrA*) gene - 40 bases (2131-2171) (Figure 10). There were around 313 new bases that followed *arrA* downstream on the 3'end. These 313 bases were the first results for *arrB* gene sequence from MLS10.



8-5-04 Template DNA 80ng - F3/R3; Primer 34JSF3

TNNNGNGTNNNAAANNNNNCCCTCACAAANTNCCGTGCGCANCATGCGAAGTGCAGTCNCCTAATTTCGCTGATCTTCAATG
CANCACCGNTTCAAACCTGCAATCGTGAAGCGCGATCAGCAAATACAAGTTGGTACCAGGAATTTAAGGATATCGACCTT
GGTGATGAAGCATGGGATGATGTGGCAAAATTGAATCAAAGGATGCTGCAGAACTTGGNNNNNNNNNACGGTGATATGGT
CCGTTTGGTTACGCCTGAAGGTCAAATCGAAGTCAAAGCCAAGCTCTGGGAAGGTACACGTCCAGGAGTTGTTGCCAAAT
GTTACGGACAAGGTCACTGGGCATATGGCCATATCGCTTCACTTGATCGTGCCTGTCAGGTCAAATCGCCAGAGGCGGGGAACAA
CAACATTATCCTCGCTCCTGTGCACGAAGCNTTAAGCGGAAGCGGGTGCAAGACACGGTGGACAGACCCGAGTNAGAGTA
GAGAAAGTA**TAA**GTTAGAGACATTATGGCGGAGGTGAACAA**ATGGCTTAAGAAAACTNTATGCANTTGACGATNGGATT**
NCAGGCGTGTTCGGCTGTGTCNGCTGNGCGGGCACTTTGTTNAAACGAGNACAATNCAGCCNACGGTNTTCNATTNG
GNTGCNCTATNNAANAAANGGANTCTGNGNCGGTTNCGGAANATCCNATAACNAINTCATNCCCACCCCTNTGAAATTN
CTGTNNTNTNNCCATGNGNNTTGGCTTCCCNNGCNTNCNNTTTTAANTNAGANNNGNTAAGNNCGNTTNCNCTNNTCG
TTTTG

805 bases

>[gi|33466103|gb|AY283639.1](#) Bacillus selenitireducens strain MLS10 respiratory arsenate reductase catalytic subunit (arrA) gene, partial cds Length=2424

Query	64	TCCGCTGATCTTTCATGGCANCACCGNITCAA	AACTGCAATCGTGAAGCGCGATCAGCAA	123	
Sbjct	2001	TCCGCTGATCTTTCATGG	AACACCG-TTCAA	AACTG-AATCGTGAAGCGCGATCAGCAA	2057
Query	124	TACAAGTTGGTACCAGGAATTTAAGGATATCGACCTTGGTGATGAAGCATGGGATGATGT		183	
Sbjct	2058	TACAAGTTGGTACCAGGAATTTAAGGATATCGACCTTGGTGATGAAGCATGGGATGATGT		2117	
Query	184	GGCAAAATTGAATCCAAAGGATGCTGCAGAACTTnnnnnnnnnn	ACGGTGATATGGTCCG	243	
Sbjct	2118	GGCAAAATTGAATCCAAAGGATGCTGCAGAACTTGGTATTCAA	ACGGTGATATGGTCCG	2177	

```

Query 244   TTTGGTTACGCCTGAAGGTCAAATCGAAGTCAAAGCCAAGCTCTGGGAAGGTACACGTCC 303
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 2178   TTTGGTTACGCCTGAAGGTCAAATCGAAGTCAAAGCCAAGCTCTGGGAAGGTACACGTCC 2237

Query 304   AGGAGTTGTTGCCAAATGTTACGGACAAGGTCACTGGGCATATGGCCATATCGCTTCACT 363
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 2238   AGGAGTTGTTGCCAAATGTTACGGACAAGGTCACTGGGCATATGGCCATATCGCTTCACT 2297

Query 364   TGATCGTCGCCGTCAAATCGCCAGAGGCGGGGAACAACAACATTATCCTCGCTCCTGTGC 423
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct 2298   TGATCGTCGCCGTCAAATCGCCAGAGGC-GGGAACAACAACATTATCCTCGCTCCTGTGC 2356

Query 424   ACGAAGCNTTAAGCGGAAGCGGGTGCAAGACACGGTGGACAGACCCGAGTNAGAGTAGAG 483
           ||||||| |||||||||||||||||||||||||||||||||||
Sbjct 2357   ACGAAGCATTAAGCGGAAGC-GGTGCAAGACACGGTGGACAGACCCGAGTAAGAGTAGAG 2415

Query 484   AAAGTATAA 492
           ||||||||
Sbjct 2416   AAAGTATAA 2424

>gi|51871105|gb|AY660885.1| Bacillus arseniciselenatis arsenate respiratory reductase
(arrA) gene, complete cds Length=2314

Score = 50.1 bits (25), Expect = 0.010
Identities = 37/41 (90%), Gaps = 0/41 (0%)
Strand=Plus/Plus

Query 160   TGGTGATGAAGCATGGGATGATGTGGCAAAATTGAATCCAA 200
           |||||||| || |||||||||||||||||||| | |||||||
Sbjct 2131   TGGTGATGAGGCGTGGGATGATGTGGCAAAATTAATCCAA 2171

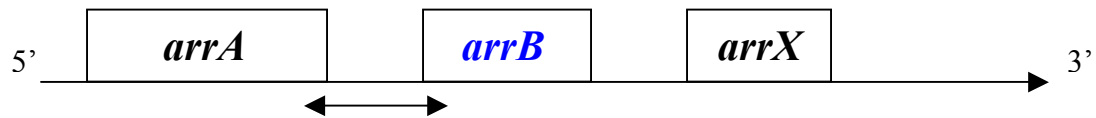
```

Figure 10: Blast results for sequence obtained utilizing 80ng template DNA – 34JSF3/ArrBDhR3 and forward primer 34JSF3 where the query is the sequence supplied for the BLAST and the subject is the sequence/s that aligned the best from GenBank. (The stop codon for the *arrA* gene - TAA is marked in red. ATG is where *arrB* gene begins. New sequence is marked in blue. *arrA* and *arrB* are separated by an intergenic sequence of 29 bases)

On alignment with earlier results from *arrB* sequencing in the lab, the new sequence diverged after ~ 40 bases and included many non-readable stretches. It was concluded from this analysis that the sequencing reaction had to be cleaner in order to obtain dependable sequence. Further optimization was carried out by altering the template concentration (B. Porter, personal communications).

The same PCR product was amplified again, the product separated on a 0.8% agarose gel, eluted, purified with the CENTRI-SEP columns and two reactions were sequenced. This time two concentrations of template DNA, 97ng and 48.5ng with the

same forward primer 34JSF3 were used. Around 680 bases were readable from the sequencing results. A figure illustrating the results from the sequencing reaction with 97ng of DNA template is shown in Figure 11.



8-5-04 Template DNA 97ng - F3/R3; Primer 34JSF3

```
TCCNGATATCCNTCGCCACAANCGAAGTCAGTCANCNTAATTTCCGCTGATCTTCNTGGCAACACCGCTTCAAANCTGC
AATCGTGAAGCGCGATCAGCAAATACAAGTTGGTACCAGGAATTTAAGGATATCGACCTTGGTGATGAAGCATGGGATGA
TGTGGCAAAATTGAATCCAAAGGATGCTGCAGAACTTGGTATTCAAACGGTGATATGGTCCGTTTGGTTACGCCTGAAG
GTCAAATCGAAGTCAAAGCCAAGCTCTGGGAAGGTACACGTCAGGAGTTGTTGCCAAATGTTACGGACAAGGTCACTGG
GCATATGGCCATATCGCTTCACTTGATCGTCGCCGTCAAATCGCCAGAGGCGGGAACAACAACATTATCCTCGCTCCTGT
GCACGAAGCATTAAAGCGGAAGCGGTGCAAGACACGGTGGACAGACCCGAGTAAGAGTAGAGAAAGTATAGTTAGAGACA
TTATGCGGAGGTGAAACAAATCGGCTAAGAAAAATTATGCAATGACGATAGATTACAGGCGTGTATCGGCTGTGCCGGCT
GTGCGGTCACTTGTAAGAACGAGAACAAATACAGACGACNCGGTCANTTGGATGCNCTNTACAAAAGGANTCCGGACNGT
TTCCGAATATCCGATACNATTNCNTGCCACCCNTTNGTAATCCCTGNGATCATGCNCCNTGTGTCNTGGGCG 712 bases.
```

>gi|33466103|gb|AY283639.1 Bacillus selenitireducens strain MLS10 respiratory arsenate reductase catalytic subunit (arrA) gene, partial cds Length=2424
Score = 787 bits (397), Expect = 0.0
Identities = 422/427 (98%), Gaps = 3/427 (0%)
Strand=Plus/Plus

Query	44	TCCGCTGATCTTCNTGGCAACACCGCTTCAAANCTGCAATCGTGAAGCGCGATCAGCAA	103
Sbjct	2001	TCCGCTGATCTTCATGG-AACACCG-TTCAAACATG-AATCGTGAAGCGCGATCAGCAA	2057
Query	104	TACAAGTTGGTACCAGGAATTTAAGGATATCGACCTTGGTGATGAAGCATGGGATGATGT	163
Sbjct	2058	TACAAGTTGGTACCAGGAATTTAAGGATATCGACCTTGGTGATGAAGCATGGGATGATGT	2117
Query	164	GGCAAAATTGAATCCAAAGGATGCTGCAGAACTTGGTATTCAAACGGTGATATGGTCCG	223
Sbjct	2118	GGCAAAATTGAATCCAAAGGATGCTGCAGAACTTGGTATTCAAACGGTGATATGGTCCG	2177
Query	224	TTTGGTTACGCCTGAAGGTCAAATCGAAGTCAAAGCCAAGCTCTGGGAAGGTACACGTCC	283
Sbjct	2178	TTTGGTTACGCCTGAAGGTCAAATCGAAGTCAAAGCCAAGCTCTGGGAAGGTACACGTCC	2237
Query	284	AGGAGTTGTTGCCAAATGTTACGGACAAGGTCACCTGGGCATATGGCCATATCGCTTCACT	343
Sbjct	2238	AGGAGTTGTTGCCAAATGTTACGGACAAGGTCACCTGGGCATATGGCCATATCGCTTCACT	2297
Query	344	TGATCGTCGCCGTCAAATCGCCAGAGGCGGGAACAACAACATTATCCTCGCTCCTGTGCA	403
Sbjct	2298	TGATCGTCGCCGTCAAATCGCCAGAGGCGGGAACAACAACATTATCCTCGCTCCTGTGCA	2357
Query	404	CGAAGCATTAAAGCGGAAGCGGTGCAAGACACGGTGGACAGACCCGAGTAAGAGTAGAGAA	463
Sbjct	2358	CGAAGCATTAAAGCGGAAGCGGTGCAAGACACGGTGGACAGACCCGAGTAAGAGTAGAGAA	2417
Query	464	AGTATAA	470
Sbjct	2418	AGTATAA	2424


```

>gi|51871105|gb|AY660885.1| Bacillus arseniciselenatis arsenate respiratory reductase
(arrA) gene, complete cds Length=2314
Score = 50.1 bits (25), Expect = 0.008
Identities = 37/41 (90%), Gaps = 0/41 (0%)
Strand=Plus/Plus

Query 140 TGGTGATGAAGCATGGGATGATGTGGCAAAATTGAATCCAA 180
          |||||      || |||||
Sbjct 2131 TGGTGATGAGGCGTGGGATGATGTGGCAAAATTAATCCAA 2171

```

Figure 11: Blast results for sequence obtained utilizing 97ng template DNA – 34JSF3/ArrBDhR3 and forward primer 34JSF3 where the query is the sequence supplied for the BLAST and the subject is the sequence/s that aligned the best from GenBank. (The stop codon for the *arrA* gene - TAA is marked in red. ATG is where *arrB* gene begins. New sequence is marked in blue. *arrA* and *arrB* are separated by an intergenic sequence of 29 bases)

Similar results as above were obtained with 48.5ng of template and the 34JSF3 primer. Sequences obtained from 3 reactions that produced results, and their alignment (Figure 12), gave a good idea that indeed the operon of interest, the '*arr*' operon, from MLS10 genomic DNA was being amplified and also, that the forward primer 34JSF3 was a good candidate for sequencing further samples.

```

F3C      GTTAGAGACATTATG-CGGAGGTGAAACAAATGGCT-AAGAAAAA-TTATGCAAT-GACG 56
F3B      GTTAGAGACATTATG-CGGAGGTGAAACAAATGGCT-AAGAAAAA-TTATGCAAT-GACG 55
F3A      GTTAGAGACATTATGGCGGAGGTGNAACAAATGGCTTAAGAAAACNTATGCANTTGACG 60
          *****
          *****

F3C      ATAGATTT-ACAGGCGTGTATC-GGCTGTGCCGGCTGTGCGGTCACTT-GTAAAAACGAG 113
F3B      ATAGATTTTACAGGCGTGTATC-GGCTGTGCCGGCTGTGCGGTCCCTT-GTAAAA-CGAG 112
F3A      ATNGGATTTNCAGGCGTGTTCGGCTGTGTGCTGNGCTGNGCGGGCACTTGTTNAAACGAG 120
          ** *
          *****

F3C      AACAAATACAGAC-GACNNGGT--CANTTGGATGCNCTNTACAAAAA---GGANTCCGGAC 167
F3B      ANCAATACAGAC-GACGGTGT--CATT-GGATGCCCTNTACAAAAA---GGAATCCGGAC 166
F3A      NACAATNCAGCCNACGGTNTTCNATTNGGNTGCNCTATNNAAANAAANGGANTCTGGNN 180
          **** *
          *****

F3C      NGTTTCCGAATA--TCCGATACNATTNCNTGCC-ACCCNTTNGTAATCCCTGNGATCATG 224
F3B      NGTTTCCGAATA--TCCGATACGATTTCTNTGCC-ANCCCTTTGTATTCACTGNGATCATG 223
F3A      CGGTTNCGGAANATNCCNATATNTCATNCCCACCCCTNTGAATTNCTGTNNTN-TN 239
          * * * *
          *****

F3C      CNCCNTGTGTCNTGGGCG----- 242
F3B      CNNCATGTGTCTGGNGNCC----- 244
F3A      NNCCATGNGNNTTGGCTTCCCNNGNCT 267
          * * * *
          *****

```

Figure 12: *arrB* sequence obtained with 80ng DNA template (F3A), 97ng (F3B) & 48.5ng (F3C) followed by Clustal W (1.82) alignment of the 3 sequences (* indicates identical bases and the non-identical bases are shown green).

Each sequencing reaction gave ~400 bases on the 3' end of *arrA* and a little over 200 continuous readable bases downstream of *arrA* gene. No further significant progress could be made with the existing primers.

The use of 34JSF4 both for sequencing and cloning had not been thoroughly probed earlier. This was a forward primer downstream of 34JSF3 and should give more sequence into *arrB* than the 34JSF3 primer. The focus was now placed on this forward primer and PCR reactions with 34JSF4 and the 4 reverse primers (Table 1, 4-8) were sought. A new fragment, ~1.1 kb amplified with 34JSF4 and ArrBDhR3 (Figure 28, Panel A) was cloned into pCR 4 TOPO and plated onto LB agar plates containing ampicillin 100ug ml⁻¹. Subsequently, selective colonies were picked and subjected to the cracking procedure and digested with EcoR1, to test for the presence of the insert. The PCR product of interest had not been inserted into the vector (Figure 28, Panel B & C).

A final attempt was made to use the existing reverse primers and to clone the 1.1 kb fragment (Figure 28, Panel A) using the same kit but plating the transformants onto LB agar plates containing 50ug ml⁻¹ kanamycin and ampicillin 100ug ml⁻¹ separately. Selected colonies were picked and analyzed as earlier. An added protocol from the manufacturer's manual for analyzing positive transformants was also performed. This included PCR reactions with primers in the kit and running them on a gel. The results from this protocol produced sequence data which matched with sequence data obtained earlier in the laboratory; therefore further research was done using these clones and sequencing the pure plasmid DNA (Figure 28, Panel B & C)

Sequencing from plasmid DNA with the use of the reverse primer 34JSF4 gave excellent results with long and clean readable sequence. The sequences also aligned with the *arrB* sequence from earlier research in the lab.

When blasted, none of the recently obtained sequences that were thought to be a part of the *arrB* gene sequence matched with *Bacillus selenitireducens* or any other closely related bacteria. Instead, all the hits for every sequence obtained were parts of cloning vectors.

On analyzing the 34JSF4 primer it was found to exactly match the vector sequence from the kit that was used for cloning. The 34JSF4 primer had been designed ignorantly by mistake, based on vector sequence that was added to the *arrB* gene sequence obtained at the same time, earlier in the lab.

At this point of time in the study, new MLS10 sequence specific forward primers and degenerate reverse primers (half degenerate primers) were designed to amplify *arrB* with a new perspective altogether. This was initiated by design of new primers with profound guidance from Professor Mahmoud Berekaa.

Sequencing the *arrB* gene from MLS10

New primers were designed for the amplification of *arrB* from MLS10, the first goal of the study. There had been fair success at amplifying the 3' end of *arrA* gene and ~250 bases at the 5' of *arrB*, the gene of interest. Results from the use of new primers should corroborate earlier findings and lead further towards achieving the goals that were set for this study.

A very rigorous protocol was followed where every detail was given utmost scrutiny. There were on one hand, three, MLS10 sequence specific forward primers whose melting temperatures ranged between 47.6°C and 50.7°C; and on the other hand a set of four degenerate reverse primers with melting temperatures ranging from 50.4°C – 62.0 °C (Table 1, 9-15). The task of optimizing the PCR reactions with respects to the size of the expected fragment (extension at 72°C) and the optimum annealing temperature based on the T_m of the primer pair involved, asked for several PCR trials in the order of almost 17 – 20 reactions.

The primer pair MArrBF1/DArrBR1 was omitted since the T_m values were 15°C apart and optimization would be futile. All other primer pair combinations were considered good candidates for amplification with the expected fragment ranging between 375 bp – 1200 bp (Table 2). After seven PCR trials the annealing temperatures for primer pairs MArrBF1/DArrBR2, MArrBF2/DArrBR1, MArrBF3/DArrBR1 and MArrBF3/DArrBR2 were successful (Table 2). The expected fragments were good candidates for sequencing as they were in agreement with expected assumptions based on primer design. Another 6 PCR trials helped optimize the annealing temperatures for the

primer pairs, MArrBF1/DArrBR4, MArrBF3/DArrBR3 and MArrBF2/DArrBR3 (Table 2) (Figure 29).

After looking at several gels for PCR reactions from different trials of annealing temperatures for the primer pairs MArrBF2/DArrBR3 and MArrBF3/DArrBR3 (Figure 29), an unusual phenomenon was observed. Each time, a dominant band would show up along with the expected fragment as a faint band. It was suggested that the reverse primer DArrBR3 is binding to genomic DNA not at the expected region but further downstream. Thus, the difference between the dominant and the expected band (~610 bases) should give us exactly, the position at which DArrBR3 consistently binds during every PCR reaction. That would mean that the MArrBF1/DArrBR3 primer pair also supports this phenomenon.

It was realized from looking back into the earlier gels that indeed this was a consistent phenomenon. An added thrill was the fact, that if indeed DArrBR3 binds further downstream, then it should be possible to sequence the PCR products using this primer in particular and obtain all of *arrB* gene sequence and even little further downstream according to the predictions. To confirm this, a PCR reaction was made with the MArrBF2/DArrBR3 fragment as template DNA and the MArrBF3/DArrBR2 primer pair at 51°C. A 520 bp fragment, the dominant and only band from the reaction was separated on the gel (Figure 30, lane 1).

Primer pair	T _m value	Expected fragment size (dominant fragment size)	Annealing temperature
MArrBF1/DArrBR2	47.6°C/55.6°C	~800 bp	50°C/52°C
MArrBF1/DArrBR3	47.6°C/57.5°C	~888 bp (~1500 bp)	_____
MArrBF1/DArrBR4	47.6°C/50.4°C	~1190 bp	53°C
MArrBF2/DArrBR1	52.3°C/62.0°C	~556 bp	56.5°C
MArrBF2/DArrBR2	52.3°C/55.6°C	~701 bp	_____
MArrBF2/DArrBR3	52.3°C/57.5°C	~789 bp (~1400 bp)	51°C
MArrBF2/DArrBR4	52.3°C/50.4°C	~1091 bp	45°C
MArrBF3/DArrBR1	50.7°C/62.0°C	~375 bp	56.5°C
MArrBF3/DArrBR2	50.7°C/55.6°C	~520 bp	51°C
MArrBF3/DArrBR3	50.7°C/57.5°C	~608 bp (~1200 bp)	51°C
MArrBF3/DArrBR4	50.7°C/50.4°C	~910/1100 bp	52.5°C - 42.5°C Touch down PCR

Table 2: Primer pairs used for amplification of *arrB* from MLS10, T_m values, expected fragment size and the optimal annealing temperature for PCR. Corrected fragment sizes are in parentheses, ____ indicates no optimal annealing temperature.

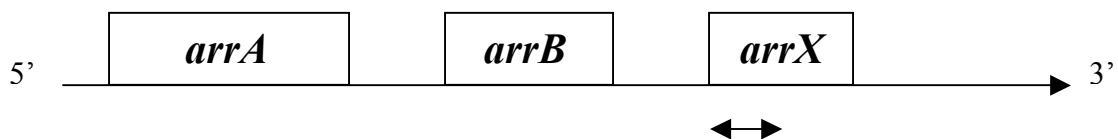
Sequence analysis

The entire *arrB* sequence was obtained after combining sequence data from several sequencing results. The fragments that were sequenced were the 1200 bp MArrBF3/DArrBR3 fragment (Figure 30), which provided 273 bases of *arrA* 5' from MLS10 followed by 157 new bases of 3' *arrB* gene. DArrBR3 reverse primer amplified 760 clean bases, which included the rest of the *arrB* gene sequence (647 bases), and

another 113 bases down stream of *arrB*, that was designated the name *arrX* (This study) (Figure 13a).

A new primer (MarrXF1) was designed based on these 113 bases to try and sequence into the ~1200 bp fragment in an effort to sequence a putative *arrD* downstream of *arrB* if it were present in MLS10. Another 67 bases were sequenced with the new MarrXF1 forward primer with a total of 180 bases corresponding to the *arrX* gene.

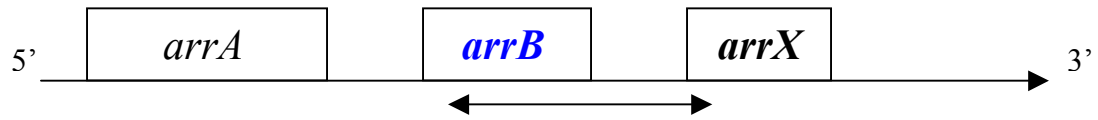
The 180 bases downstream of *arrB* was BLAST searched. The closest and only match from blastx was a hypothetical protein from GK1786 *Geobacillus kaustophilus* HTA426 with 52% identity. When this hypothetical protein was in turn BLAST searched the closest hit was another hypothetical protein Bant_01002722 from *Bacillus anthracis* str. A2012. Nothing further could be deduced from these results regarding the identity of ArrX.



TCGACTTTGAAGGACGTGATCATCATGAGTA**ATGCAGAATTAGGAGAGTTTC**CATTATAAAGCCAA
TCTCTTCAAGTATTAAGCGAATTGTATAAAGAGCCGGGTGCGGATTTACCCGGCTACATGGACTA
TTTAATCGAGGCCTTGGAAGAACTTCAACCTGCATCGGCTGCAAGAACTG

Figure 13a: *arrX* from MLS10 (The forward primer MarrXF1 is shown in **bold**)

MArrBF3/DArrBR4, ~910 bp fragment (Figure 29, Panel B) helped corroborate the data from MArrBF3/DArrBR3, but gave only 524 bases corresponding to *arrB* with the use of the DArrBR4 reverse primer (Figure 13b).



TCGGCTGTGCCGGCTNTGCGGTCANTTGTAAAAACGAGAACAATACAGACGACGGTGTCTC
AATTGGATGCACTATACAAAAAGGAATCCGGACAGTTTCCGAATATCCGATACGATTTC
ATGCCAACCCCTTTGTAATCACTGTGATCATGCGCCATGTGTCTATGGCGTGCCCGGTCACA
GCGATGTACAAAGATGAGGATGGATTAACGCTGCATGACGCAGACCGGTGTATTGGTTGT
AAAGCATGTATGACAGCATGCCCCGTATGGTGTTATTTTCCTTTAATAAAAAGGATCCGCAC
CAGTACTGGAATGAAAACAATGCATGGAGTGAGCTTTCAGCTACACCGAAAGAAGTACAA
GAAAGTGCTGGAGCAGATCTGCCATATTACAACCCGGAACGTGATTATAACTACGAAGCA
ATTCGCTATCGCGGAATCGTTGAGAAATGTCAGTTCTGTGATCACAGACTAGCTCGCGAT
GAACAGCCATACTGTGTAGAGAGATGTCCATCTGAAGCGATGTATGTCGGCGACTTGAAC
GATCCGGATGACAAAATTCATGAGCTGTTAAAATACAGTCATGAAGGGTTAAAAGAAGAG
CTTGGAACGAAGCCAAAAGTCTTCTACTTGCAGAAAGTTCAGTAAATAATCGACTTTGAAG
GACGTGATCATCATGAGTAATGCAGAATTAGGAGAGTTCCATTATAAAGCCAATCTCTTC
AAAGTATTAAGCGAATTGTATAAAGAGCCGGGTGCGGATTT 760 bases

Figure 13b: MLS10 Whole clone sequenced with the DArrBR3 reverse primer. (647 bases in blue correspond to *arrB* gene followed by 113 bases downstream)



GGGAAAGGTACANGTNCCAGGAGTTGTNNCCAAANNTACGGACAAGGTCACTGGNCATAT
GGCCATATCGCTTCACTTGATNGTCGCCGTCAAATCGCCAGAGGCGGGAACAACAACATT
ATCCTCGCTCCTGTGCACGAAGCATTAAGCGGAAGCGGTGCAAGACACGGTGGACAGACC
CGAGTAAGAGTAGAGAAAGTATAAGTTAGAGACATTATGCGGAGGTGACACAAATGGCTA
AGAAAAATTANNCAATGACGATAGATTTACAGGCGTGTATCGGCTGTGCCGGCTGTGCGG
TCACTTGTAAAAACGAGAACAATACAGACGACGGTGTCAATTGGATGCACTATACAAAAA
AGGAATCCGGACAGTTTCCGAATATCCGATACGATTTCATGCCAACCCCTTTGTAATCACT
GTGATCATGCGCCATGTGTCTATGGCGTGCCCGGTACAGCGATGTACAAAGATGAGGATG
GATTAACGCTGCATGACGCAGACCGGTGTATTGGTTGTAAAGCATGTATGACAGCATGCC
CGTATGGTGTTATTTTCCTTTAATAAAAAGGATCCGCACCAGTACTGGAATGAAAACAATG
CATGGAGTGAGCTTTCAGCTACACCGAAAGAAGTACAAGAAAGTGCTGGAGCAGATCTGC
CATATTACAACCCGGAACGTGATTATAATTACGAAGCAATTCGCTATCGCGGAATCGTTG
AGAAATGTCAGTTCTGTGATCACAGACTAGCTCGCGA 757 bases

Figure 13c: MLS10 whole clone sequenced with DArrBR4 reverse primer. (524 bases in blue correspond to MLS10 *arrB* gene, 233 bases in black = MLS10 *arrA* gene 3')

Another dominant band (~1.1 kb) that resulted from the PCR reaction with the MArrBF3/DArrBR4 primer pair was also sequenced with the DArrBR4 primer and the results (845 bases) were clearly not a portion of the *arrB* gene (Figure 29, Panel B).

BLAST searches with these 845 bases (blastx) showed the closest match with ABC transporter (ATP-binding protein) *Bacillus halodurans* C-125 (60% identity), nitrate/sulfonate/bicarbonate ABC transporter, ATP-binding protein *Bacillus clausii* KSM-K16 (53% identity) and ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component, *Burkholderia fungorum* LB400 (52% identity). This was suggestive for the presence of an ABC transporter (ATP binding protein) in *B. selenitireducens* strain MLS10, a protein/s that is/are part of the largest family of proteins.

Identification of the arrB gene

Surprisingly enough, *arrB* from MLS10 was a small gene with 693 bases i.e. 231 amino acids. In *B. selenitireducens* strain MLS10, the ORF encoding the *arrB* gene is located immediately downstream of the ORF that encodes the *arrA* gene as in *Shewanella* sp. strain ANA-3, *W.succinogenes* and *Desulfitobacterium hafniense*. The predicted molecular mass of the *arrB* gene product is 26.3 kDa and its predicted pI is 5.50 (www.expasy.org) [Molecular mass was predicted to be 34 kDa from prior study (Afkar et al., 2003)].

arrB gene is directly downstream of the *arrA* gene (AY283639) which is 2424 bp i.e. 807 amino acids. The predicted molecular weight of ArrA is 89.9 kDa and its predicted pI is 5.11 (www.expasy.org). Similarly, from earlier studies the calculated

molecular mass of ArrA was 90.9 kDa and $pI = 5.12$ (Afkar et al., 2003). ArrA has a cysteine-rich motif (C-X₂-C-X₃-C-X₂₇-C) beginning at amino acid residue 10 and extending to position 45 that is predicted to coordinate an iron-sulfur cluster (Saltikov and Newman, 2003) and a molybdopterin dinucleotide-binding domain (residues 633 – 696) (Figure 14) .

The amino acid sequence was analyzed using the signal peptide production program SIGNALP (www.cbs.dtu.dk/services/SignalP/) to identify a Tat recognition region and was found to contain no such Tat signal sequence (RR-twin arginine) similar to what has been found in *Shewanella* sp. strain ANA-3 (Saltikov and Newman, 2003). Since the N-terminal sequence from MLS10 ArrA is still not available it cannot be ascertained for sure that the arsenate reductase enzyme in MLS10 is not secreted.

```
EQGEWIASVC QGCTAWCAVQ VYRIDGRATK VRGNPNAKAN HGHS CVRSHI GLQQVYDPDR VKQPMKRTNP
NKGREDPEF VPISWEEAMD TIADKIIELR ENNETHKFSV WRGRYTSNNG ILYGNMPKII GSPNNISHSS
ICAESKFKGR YYTERYWGGA DYDHENALYE IFWGGDPIAT NRXVPHTASI WGELESDRATL ACVDPRFSTT
AANQMNGCRL IPGEDGAIAS AIAHVILTEG VWYKPFVGF KDGQNRFEV RDVNEDDFEE IQTHGLVKWW
NLELKDKTPE WAAERSGIDA DQIYRVARGF ANAAPKAISF SSPGSSMTIR GGYTAMTQAA LNLVGSADN
LGGVISNGIS VPNNGFDPDS DYIDEIAANG LEQDRIDWGG RLEFPALKDS KSGGVKVTNT VADAVLAEDP
YEFKSRIELL DELQFLQPRN GSLGQSFGQD SIHGAYDCQS GEQTHFADIV LPVPHSQFER QSPVAGSNGN
LHRHLHLQNK VIESPFDIRV DETEIPWMIG ESLEKKGYSN LIDYFRNEFR DPETGEAPT NAAEFDEIATK
HYTHPVWDPT FEKDGQIDG WEYKRLGTW NTNKYEFQK WDGNGWTETG QFEFYSETLK VALQDHADKH
NASIDDVMEA TFNTAKGELA FVPHYEPAMR VGDESEYPLI FMEHRSKLNR EARSANTSWY QEFKDIDLGD
EAWDDVAKLN PKDAAELGIQ NGDMVRLVTP EGQIEVKAKL WEGTRPGVVA KCYGQGHWAY GHASLDLRRR
QIARGGNINI ILAPVHEALS GSGARHGGQT RVRVEKV 807 AA
```

CD-Length = 111 residues, only 55.0% aligned
Score = 63.8 bits (155), Expect = 8e-11

```
Query: 633 DLGDEAWDDVAKLNPKDAAELGIQNGDMVRLVTPEGQIEVKAKLWEGTRPGVVAKCYGQG 692
Sbjct: 23 MLYKVRPRPFVEINPEDAAKLGKDGDLVRVFNRRGEVVARAVVTDRIIPGVV---FIPG 79

Query: 693 HWAY 696
Sbjct: 80 GWD 83
```

Figure 14: MLS10 ArrA protein sequence and conserved domain.

Above: the amino acid sequence of MLS10 ArrA- 807 residues, showing the cysteine-rich motif (C-X₂-C-X₃-C-X₂₇-C), the cysteine residues are shown red.

Below: Molybdopterin dinucleotide binding domain, residues 633-696. Identical residues are shown red and similar residues are shown blue.

The sequence analysis of ArrB indicates that it could contain four iron-sulfur clusters at the following cysteine residues: (i) cys 15, 18, 21, 25; (ii) cys 60, 63, 68, 72; (iii) cys 92, 95, 98, 102; and (iv) cys 165, 168, 180, 184. The first and third iron-sulfur clusters share the binding motif (C-X₂-C-X₂-C-X₃-C) while the second and fourth clusters are different with a (C-X₂-C-X₄-C-X₃-C) and (C-X₂-C-X₁₁-C-X₃-C) motif respectively. Ribosome binding site (RBS) occurs 13 bp upstream of *arrB* (GAGG) (Figure 15).

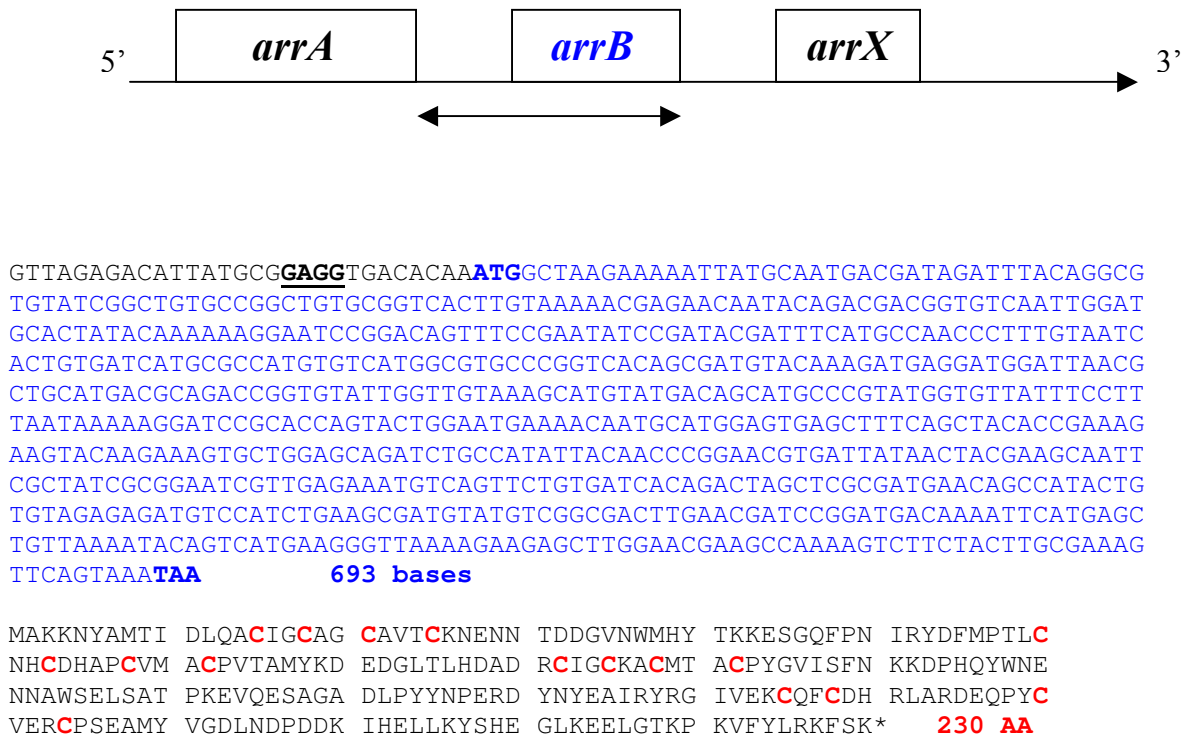


Figure 15: MLS10 DNA and protein sequence (ArrB). Above :The *arrB* gene –693 bases is shown in blue with **ATG** the start codon and **TAA** the stop codon in **bold**, RBS is underlined. Below: the amino acid sequence –230 residues corresponding to the above 693 bases excluding * - the stop codon. All the cysteine residues are shown in red.

ArrB from MLS10 also does not have a Tat signal sequence and on analyzing the predicted protein ArrB with SIGNALP (www.cbs.dtu.dk/services/SignalP/) this was further clarified. BLAST searches with *B. selenitireducens* strain MLS10 ArrB showed the closest match with molybdopterin oxidoreductase (HMOA) from *Wolinella succinogenes* (51%) and ArrB from *Shewanella* sp. ANA-3 (50%) followed by *Desulfitobacterium hafniense* DCB-2, iron-sulfur containing dehydrogenase (37%).

MLS10 ArrB alignment with the above-mentioned matches is as follows. The identity and similarity is calculated based on the number of asterisks (*) indicating identity and dots (. & :) indicating similarity.

CLUSTAL W (1.82) multiple sequence alignment

```

ArrBShew      ---MRLGMVIDLQKCVGCGGCSLACKTENNTNDGIHWSHHIATTEGTFPDVKYTYIPTLC 57
WolArrB       -MAKNYAMVIDLNRCVGCGACDLACKSENNTPVGIHWSNHVVETTGVPNVGFTYRPTLC 59
MLS10ArrB     MAKKNYAMTIDLQACIGCAGCAVTCNENNTDDGVNWMHYTKKESGQFPNIRYDFMPTLC 60
               .  .*.***: *:*..*  ::*..***  *::*  ::      *  **:: : :  ****

ArrBShew      NHCDDAPCVKVCPPTGAMHKDKR-GLTLQNNDECIGCKKCMNACPYGVISFNAATPHRRWQ 116
WolArrB       NHCDDAPCVRACPPTGAMYKDKEGSMTLHNADKICIGCKSCMLADPYHVIYNSKQPHYKWT 119
MLS10ArrB     NHCDHAPCVMACPVTAMYKDED-GLTLHDADRCIGCKACMTACPYGVISFNKKDPHQYWN 119
               ****.***. .**..**:*:  *::*:  *.*****  *  *  *  *  *  *  *  *  *

ArrBShew      DDSEVVANGTVSPLMLLKRTGATATPNENPERGDTYPMIRPKRTTEKCTFCDHRLDKGLN 176
WolArrB       TTKPTVM-GMFSPEKEMADKSGAP-YPYFNVERGKTYEALRFKGLVEKCSFCDHRVKAGLM 177
MLS10ArrB     ENN-AWSELSATPKEVQESAGAD-LPYYNPERDYNYEAIRYRGIVEKCFQCDHRLARDEQ 177
               . .      :*  : . :**  *  *  *.  .  *  :  :  .***  *****:  .

ArrBShew      PACVDA CPSEARVIGDLDDPQSKVSQLIKHLKPMQLKPEAGTGPRVFYIRSFGVKTAY 234
WolArrB       PHCVVSCPGDARIFGDLNDPKSSVSQLLAQYPAETLLPHKGTKAKVFYIRSYARA--- 232
MLS10ArrB     PYCVERCPSEAMYVGDLNDPDDKIELLK-YSHEGLKEELGTPKPKVFYLRKFSK---- 230
               *  *  *  *  *  :  *  .***:*..*:  :*:  :      *  .  *  *  :***:*..:

```

Figure 16: Protein sequence alignment - ArrB from MLS10, *Wolinella succinogenes* and *Shewanella* sp. ANA-3 ([*] indicates amino acid identity , [. & :] indicate similarity, conserved cysteine residues are shown red)

ArrB from MLS10 has 40% identity and 55% similarity with ArrB from the three organisms listed in figure 16. Pair wise alignment increases the identity and similarity, thus showing that the arsenate reductases are a separate class of enzymes and have conserved domains.

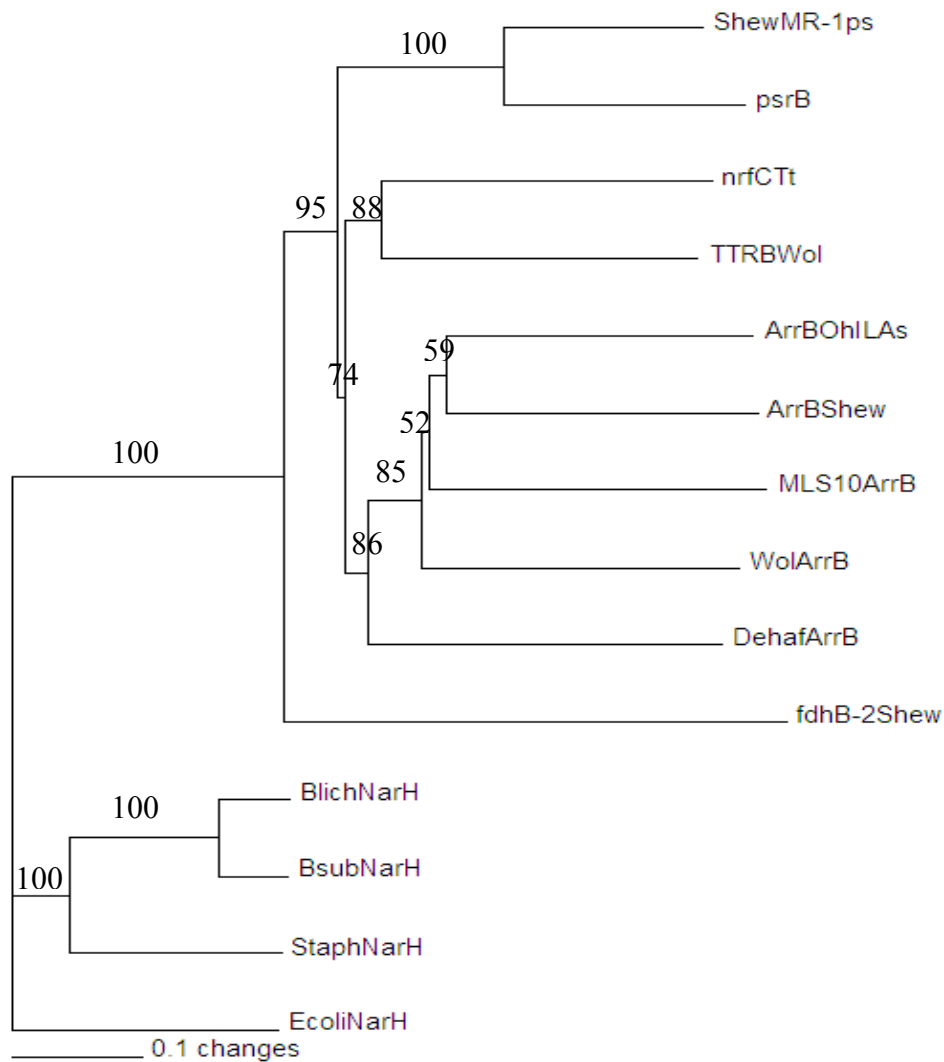


Figure 17: Neighbor joining, rooted tree with bootstrap values generated for protein sequences with significant alignment from the BLAST search for MLS10 ArrB.

The phylogenetic tree – neighbor joining, rooted with *E.coli* NarH as the outgroup generated with protein sequences that showed significant alignment in blastp suggests that MLS10 ArrB is similar to that from *W.succinogenes* (Figure 17).

The amino acid composition of MLS10 *arrB* gene product shows that the protein is composed of 34.4% non-polar residues and an equal number of polar residues of 35.2% (Table 3). The hydropathy plot (Kyte and Doolittle) calculated with an average window size of 9 amino acids indicates several regions that are hydrophilic and hydrophobic in nature. There are atleast 8 negative hydrophilic peaks above a threshold of -2.0 , 4 each extending between residues 20 and 64 and 107-226 respectively. Six other hydrophilic peaks above -1.0 are also seen. The hydrophobicity is weak with only 6 peaks above a $+1$ extending throughout the first half of the protein i.e. ~ 110 residues from the N-terminus (Figure 18). Altering the average window size between 5 to 25 amino acids did not affect the hydropathy profile of the protein.

From earlier studies the cellular location of the arsenate reductase from *B.selenitireducens* has been suggested to be membrane-bound (Afkar et al., 2003). The pure Arr from MLS10 after gel filtration chromatography generated 2 bands corresponding to a 110 kDa ArrA and a 34 kDa ArrB indicating that the active enzyme is an ArrA/ArrB complex (Mirunalni Thangavelu, Masters thesis 2004), whose activity was equal in the whole cell lysate and the membrane fraction.

The Tat pathway functions to transport folded proteins especially those containing redox cofactors ex. Molybdenum and Fe-S clusters in MLS10, involved in respiratory electron transport chains across the cytoplasmic membrane. No Tat signal peptide was

detected in the intergenic 29 bases between *arrA* and *arrB* genes in MLS10. No promoters or regulatory regions similar to those in *Bacillus subtilis narGHJI* (the nitrate reductase operon) were detected. 5' end sequencing of the *arrA* gene and the upstream region will help in detecting a Tat signal sequence S/T-R-R-X-F-L-K if present, similar to that found in the N terminus of ArrA from *Shewanella* ANA-3 and a similar putative signal found in *D.hafniense*. This would be suggestive of the enzyme being exported across the cytoplasmic membrane wherein it becomes membrane bound and linked to the electron transport chain involved in energy generation. The hydrophilic residues would facilitate in this transport in the cytoplasm.

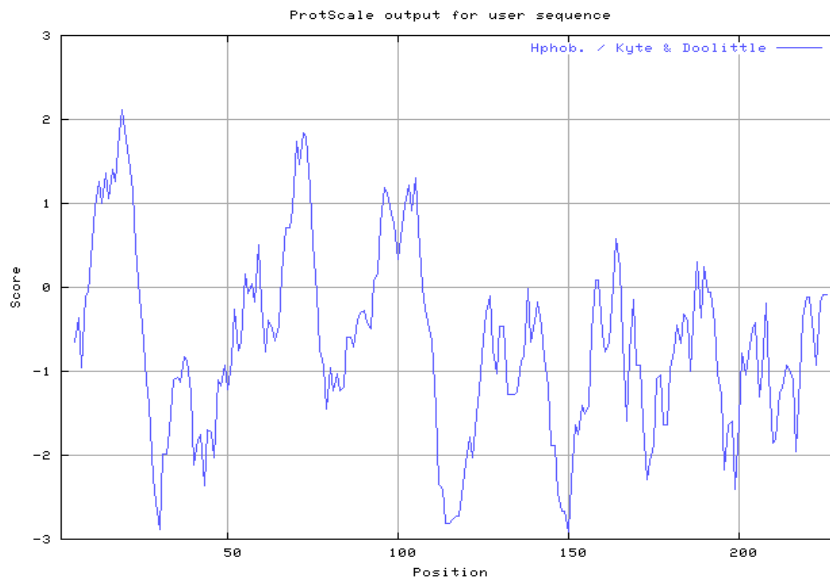


Figure 18: Hydropathy profile of the MLS10 ArrB protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis.

MLS10 ArrB		
	No.	%
Non polar		
Ala (A)	18	7.8
Val (V)	10	4.3
Leu (L)	13	5.7
Ile (I)	8	3.5
Pro (P)	13	5.7
Met (M)	8	3.5
Phe (F)	6	2.6
Trp (W)	3	1.3
Total	79	34.4
Polar		
Gly (G)	12	5.2
Ser (S)	8	3.5
Thr (T)	10	4.3
Cys (C)	16	7.0
Tyr (Y)	15	6.5
Asn (N)	14	6.1
Gln (Q)	6	2.6
Total	81	35.2
Acidic		
Asp (D)	18	7.8
Glu (E)	17	7.4
Total	35	15.2
Basic		
Lys (K)	18	7.8
Arg (R)	9	3.9
His (H)	8	3.5
Total	35	15.2
Total residues	230	
Molecular weight	26.3kDa	
Estimated pI	5.5	

Table 3: Amino acid composition of the predicted product of the *arrB* gene from MLS10

***Clostridium* sp. strain OhILAs - *arrA* and *arrB* genes**

The 19 primers (Table 1, #17-35, Figures 8 & 9) were optimized for PCR, to produce 4 fragments for sequencing the *arrA* gene and 3 fragments to sequence the *arrB* gene from OhILAs. The optimal annealing temperatures for the primer pairs, their T_m values and the expected fragments that were generated are summarized in the following table.

Primer Pair	T _m value for the primer pair	Expected fragment size	Optimal annealing temperature
arrAUF1/arrAUR2	56.2°C/53.1°C	245 bp	48.7°C
arrAUF1/arrAUR3	56.2°C/60.3°C	852 bp	54°C
arrAUF1/arrAUR5	56.2°C/48.7°C	1825 bp	54°C
arrAMLSF4/arrAUR6	56.2°C/54.9°C	348 bp	50°C
MArrBF3/DArrBR1	50.7°C/62.0°C	375 bp	56.5°C
arrAUF5/DArrBR2	48.7°C/55.6°C	1.2 kb	53°C
arrAUF5/DArrBR4	48.7°C/50.4°C	800 bp	50°C

Table 4: Primer pairs used for amplification of *arrA* and *arrB* from OhILAs, T_m values, expected fragment size and optimal annealing temperature for PCR.

The expected fragments were gel eluted, purified and sequenced. A suitable forward or reverse primer (Table 1) was used to sequence each of the above fragments. Depending on the primer, the sequencing result was either good with continuous readable sequence or the vice versa.

Sequence data from several amplicons were used to obtain the *arrA* and *arrB* genes from OhILAs (122 bases missing from *arrA*, which still need to be amplified). They were in the order as follows: Reverse primer arrAUR3 provided the first 783 bases corresponding to the 5' of *arrA* when a 852 bp fragment amplified with primers arrAUF1/arrAUR3 was sequenced (Figure 31, Panel A); reverse primer arrAUR5 provided the following 766 bases when a 1825 bp fragment amplified with primers arrAUF1/arrAUR5 and reverse primer arrAUR6 provided another 217 bases when a 348 bp fragment amplified with primers arrMLSF4/arrAUR6 was sequenced (Figure 31, Panel B); forward primer arrAUF5 provided 265 bases which were the 3' of *arrA* and the 5' of *arrB* when a 375 bp fragment amplified with primers MArrBF3/DArrBR1 was sequenced (Figure 32). The reverse primers DArrBR2 and DArrBR4 provided rest of the *arrB* gene sequence with 557 bases sequenced from a 1.2 kb fragment amplified with primers arrAUF5/DArrBR2 and 607 bases sequenced from a 800 bp fragment amplified with the primers arrAUF5/DArrBR4 respectively, with overlap in between (Figure 32).

Based on sequence alignment with the *arrAB* gene contig from MLS10 it was concluded that only around 122 bases were missing from the 3' of *arrA*. The sequence analysis was carried out using these 2956 bases (including 122 missing bases denoted by an N in nucleotide sequence and an X in protein sequence).

Sequence analysis consisted of BLAST searches similar to that done for the *arrB* gene from MLS10.

The entire 2956 bases were pasted onto the ORF finder window (www.ncbi.nlm.nih.gov/gorf) and three open reading frames were analyzed. These were at +2 (152 –988 = 836 bases), +1 (1075 –2385 = 1238 bases) and +1 (2431 –2952 = 521 bases). When BLAST searched the first +1 open reading frame showed a conserved molybdopterin dinucleotide binding domain extending between residues 339 – 425 at the 3' end. On blasting this ORF the sequences that produced significant alignment were the arsenate respiratory reductase from *C.arsenatis* 47% identity, Shewanella ArrA 42%, arsenate respiratory reductase from *B.arseniciselenatis* 43%, Molybdopterin oxidoreductase, molybdenum binding subunit from *W.succinogenes* 40%, ArrA from *B.selenitireducens* 38%, arsenate respiratory reductase from *S.barnesii* 41%, and the polysulfide reductase (psrA) and thiosulfate reductase (fdhA) from *W.succinogenes* 24% identity.

The +2 open reading frame had a molybdopterin oxidoreductase conserved domain at the 5' end and a Nitrate reductase alpha subunit (NarG) conserved domain in its 3' end. BisC, anaerobic dehydrogenase (selenocysteine containing) and FwdB, Formylmethanofuran dehydrogenase subunits B were the two other conserved domains that were indicated.

The +2 (152 –988) and +1 (1075 –2385) open reading frames corresponded to the first and second halves of the *arrA* gene. When the second ORF was BLAST searched, significant alignments to all the above sequences from the same organisms showed up, but with slightly higher identity than the other +1 ORF. The sequences that produced significant alignments were the arsenate reductase from *Bacillus arseniciselenatis* 61%, *Bacillus selenitireducens* ArrA 59%, Molybdopterin

oxidoreductase from *Wolinella succinogenes* 56%, ArrA *Shewanella* ANA-3 55%, arsenate reductase ArrA from *Chrysiogenes arsenatis* 56%, *Sulfurospirillum barnesii* SES-3 52%, *Shewanella* HAR-4 ArrA 59%, and a 28% identity to the fdhA (thiosulfate reductase) from *Wolinella succinogenes* and psrA (polysulfide reductase, subunit A) from *Shewanella oneidensis* MR-1.

From the BLAST searches it is evident that the N-terminus of OhILAs ArrA is not complete. Around 50 amino acids were added to the ArrA sequence after close analysis based on ArrA sequence from the BLAST search. These were not included by the ORF finder program and therefore were manually added to the gene sequence based on sequence similarity.

Based on this preliminary sequence data which has an incomplete 5' end and several bases missing internally, the OhILAs *arrA* gene is ~2298 nucleotides in length and the predicted gene product is ~766 amino acids long. This result is completely based on the data from the current study and needs to be validated by duplicating the sequencing protocols.

The ORF (2431 –2952) bases corresponding to the *arrB* gene was at a +1 position and when BLAST searched, the HybA, Fe-S- cluster containing hydrogenase was predicted for the entire 521 bases that were blasted. The protein blast search showed the closest matches as ArrB *Shewanella* ANA-3 50%, *Wolinella succinogenes* 40% and psr subunit B from *Shewanella oneidensis* MR-1, 38%.

The OhILAs *arrB* gene is 522 nucleotides long and the predicted gene product, the ArrB protein is 174 amino acids long, the average size of the smaller beta subunit of the arsenate reductases falling within this range.

OhILAs ArrA gene

```

-----AWCAKQVYXVDGRAIKIKANDESKIHGNDCPRAHLAVQQVYDADRIKQPMKRTN 55
: ** : * * : *** * : . * : : * * * : . : * * * . * : * *

PKKGRDEDPMFVPISWDEAMDTLADKIMELRTNNETHKFMLLRGRYTNITDLFYSSMPKI 115
* : * : . * * * * * : : * : * : * : * . * * : : * * : . : * . : :

IGSPNNISHSSICAEAEKFGPYYTQGYWNYRDYDVQNTKYMICWGVDPLCSNRQVSHYLN 175
* * * * * : * * : . * * : * . * : * . * : : : * * * : . * * . .

VFGELLKKGVKIVTIDPKYSSTAAKSHVWMPVIPGQDGALATAMAHVILAEGLWSREFVG 235
: * . . : . : * : * : * : : * * * . : * * * * * . * * : : *

DFTDGNNAFISGKNVKE-----ETFEEIHTHGVVKWNLELKDRTPEWAEAICGV 285
* * : * : * * . * * : * : * : * : * * * * * * * * * * * :

DAATIRNVAIEFAKAAPYSVMFMGGGSNMXXVGAYNSX----- 323
. * : * : . * . : . * : * * * :

-----AVPCAMDEMXKTNSKNKKMDQRGXKQFPALKEGKPGSGVITNRVADAILAE 374
: . * . : : : * * : * * * * * . * . : * . * : : :

DPYLPKVVIGYFCNFNYSCPGTDRWDKAMSKIEFFAHCVTHYSEMSHFADLLLPSTHMF 424
. * * * : : : : : : : . : . * : : * * : * * *

EQMA-SAAQKANSYTHLWITQRLIEPVYDVKNPESEVLWMLAEKLEQRGSNLLDFCKT- 474
* : . . * : : : : : : * : * : * : : * . * . : .

VKDPETGSEPTNGLEFELYAYKTLTQSIWDPTKYTELENHGDKFNSWNEFIQKGVWNSDX 524
. * * * * . : * * : * . : * . . * * . * . * . * . * . * .

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-XDDILATCKYTGKGEQAFIP 574
* : : : . : * . * : *

HYEPPYMEGKVEQYPFAYIDSKSRLAREGRSANCSWFQEFKDADPGDIKWSDCIKMNPKD 624
* * * . * . : * : : * : * : * . * : * . * * * * * . * * : * *

AIELGLKDGDKVKVTSPTGEIVTTLKLWEGLRPGTVQKNYGQGHWAMGHVAAEDFEKKLS 674
. : * : : * * : : . . : * * * * * * * * : * * * * * * * * .

RGGNSNVIIPADYERLSGATAYYG-SFRVKIERA 709
* * * . . * * * . * * * * * .

```

Figure 19: OhILAs ArrA protein alignment.([*] indicates identity and [: & .] indicates similarity)

The results for OhILAs ArrA from this current study suggests that the ArrA protein is ~766 amino acids in length and the *arrA* gene is ~2298 nucleotides in length. Based on this sequence data the predicted molecular mass is 75.2 kDa and the predicted pI is 5.86 (www.expasy.org). The amino acid sequence was analyzed using the signal peptide production program SIGNALP (www.cbs.dtu.dk/services/SignalP/), but no Tat signal was found as was expected since the N terminus is not completely sequenced. Figure 13 corroborates this result, the first two cysteine residues are missing i.e. nucleotide sequencing is incomplete at the 5' of OhILAs ArrA and also further upstream to detect the Tat signal peptide (RR). The cysteine residues in the alpha subunit of the arsenate reductases follow the conserved motif of (C-X₂-C-X₃-C-X₂₇-C), and OhILAs ArrA fulfills this criteria a 50% with the third and fourth cysteine residue at the correct position in the conserved N-terminal domain. A conserved molybdopterin dinucleotide binding domain is present at the C-terminus (Figure 14, Boxed residues in the protein sequence alignment) suggesting that the protein belongs to the DMSO reductases family of molybdoenzymes. GQGHWA is another conserved sequence in the C-terminus whose function is not known.

The protein sequence alignment showed that OhILAs ArrA shares 32% identity and 63% similarity with the ArrA sequences from *Shewanella* ANA-3, *W.succinogenes*, and *B.selenitireducens* strain MLS10 (Figure 14). The identity and similarity decreased when aligned with four additional ArrA sequences from *D.hafniense*, *B.arseniciselenatis*, *S.barnesii* and *C.arsenatis*. Based on the sequences that had significant alignment with OhILAs ArrA sequence the neighbor joining, rooted (fdhAWol), phylogenetic tree with bootstrap values was constructed using PAUP

OhILAs ArrA.1		
	No.	%
Non polar		
Ala (A)	23	8.3
Val (V)	19	6.8
Leu (L)	14	5.0
Ile (I)	17	6.1
Pro (P)	13	4.7
Met (M)	12	4.3
Phe (F)	11	4.0
Trp (W)	8	2.9
Total	117	41.1
Polar		
Gly (G)	20	7.2
Ser (S)	17	6.1
Thr (T)	16	5.8
Cys (C)	5	1.8
Tyr (Y)	12	4.3
Asn (N)	17	6.1
Gln (Q)	4	1.4
Total	91	32.7
Acidic		
Asp (D)	15	5.4
Glu (E)	17	6.1
Total	32	11.5
Basic		
Lys (K)	18	6.5
Arg (R)	10	3.6
His (H)	7	2.5
Total	35	12.6
Missing residues	3	1.1
Total residues	278	
Molecular weight	26.3 kDa	
Estimated pI	5.5	

Table 5: Amino acid composition of the predicted product of the +2 ORF (152 –988) from OhILAs corresponding to the first half of the *arrA* gene (*arrA.1*).

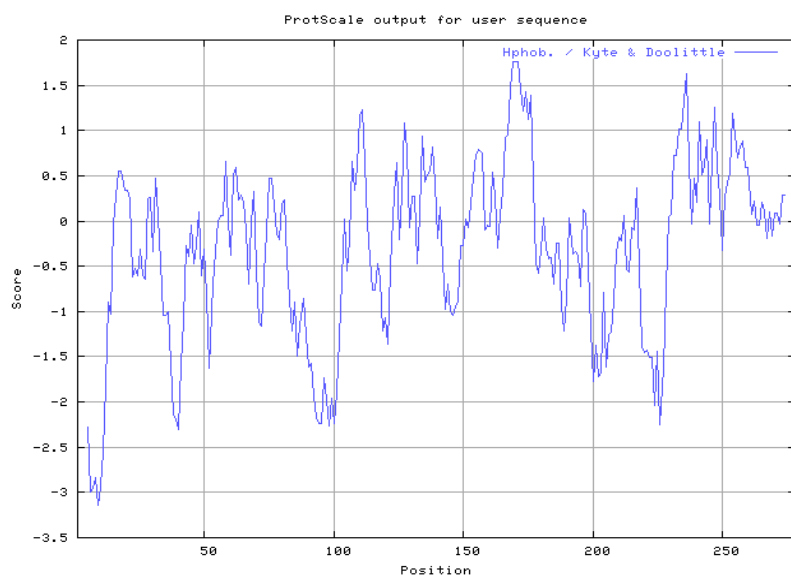


Figure 21: Hydropathy profile of the OhILAs ArrA.1 protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis.

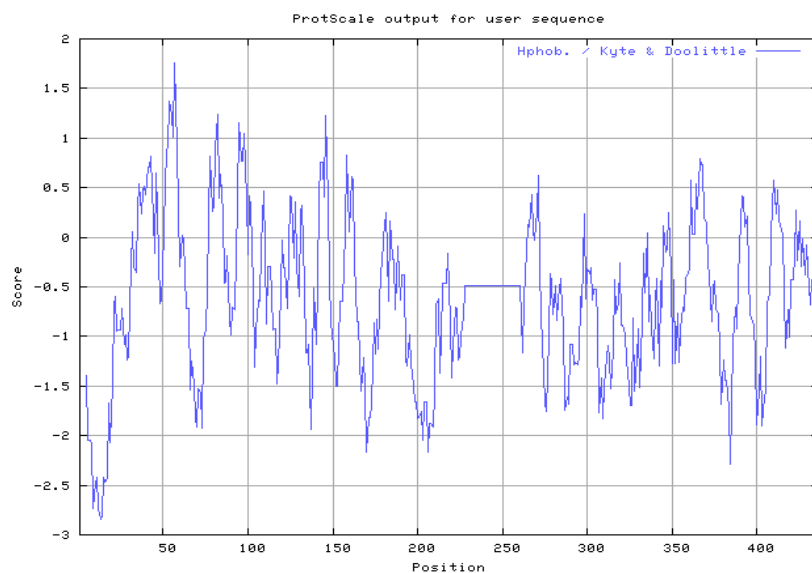


Figure 22: Hydropathy profile of the OhILAs ArrA.2 protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis.

OhILAs ArrA.2

	No.	%
Non polar		
Ala (A)	27	6.2
Val (V)	18	4.1
Leu (L)	26	6.0
Ile (I)	18	4.1
Pro (P)	20	4.6
Met (M)	11	2.5
Phe (F)	18	4.1
Trp (W)	10	2.3
Total	122	33.9
Polar		
Gly (G)	28	6.4
Ser (S)	25	5.7
Thr (T)	22	5.0
Cys (C)	7	1.6
Tyr (Y)	18	4.1
Asn (N)	18	4.1
Gln (Q)	13	3.0
Total	131	29.9
Acidic		
Asp (D)	25	5.7
Glu (E)	31	7.1
Total	56	12.8
Basic		
Lys (K)	35	8.0
Arg (R)	13	3.0
His (H)	10	2.3
Total	58	13.3
Missing residues	43	
Total residues	436	
Molecular weight	45 kDa	
Estimated pI	5.77	

Table 6: Amino acid composition of the predicted product of the +1 ORF (1075 –2385) from OhILAs corresponding to the second half of the *arrA* gene (*arrA.2*).

The hydropathy profile of the ArrA protein from OhILAs is deduced from the amino acid composition of two separate ORFs. Based on the amino acid composition of the ArrA.1 protein, non-polar hydrophobic residues are higher - 41.1% compared to a lower percentage of polar hydrophilic residues of 32.7% (Table 5). The corresponding hydropathy profile for the predicted protein shows a more hydrophilic N terminus versus a more hydrophobic C terminus. Residues extending between positions 1 – 110 and 180-220 are hydrophilic with several negative peaks. Residues 100-180 are primarily hydrophobic with several positive peaks (Figure 18).

The hydropathy profile of the ArrA.2 protein is significantly different with the amino acid composition revealing an equal percentage of hydrophobic (33.9) and hydrophilic (29.9) residues (Table 6). The hydropathy plot based on this composition reveals a hydrophilic N and C terminus, with residues 1 –200 being more negative, i.e. hydrophilic than hydrophobic. Also, around 43 residues are missing internally between 230 and 273 amino acid positions (Figure 19).

Based on this preliminary data it cannot be ascertained if the gene structure of *arrA* in OhILAs is different from those in MLS10, ANA-3 or *C.arsenatis*, which are known to be the larger subunit of the respiratory arsenate reductase. Further sequencing would provide data to fill the missing gaps in the DNA sequence and thereby also generate a predicted protein which could be composed of one subunit unlike the current results which predict 2 ORFs, each corresponding to the first and second halves of *arrA* genes from the different arsenate respiring prokaryotes.

Identification of the *arrB* gene in *OhILAs*

TATCAATAGAACGATTG**GAGG**TGTTTTAATTGGCAAAATATGGA**ATGGTAATTAACCTGGAGAA**
ATGTGCTGGATGTGCAGCTTGTTCAATTGCCTGCAAAAATGAAAAACAATGTTCCGGATGGAATAT
TCTGGTCCCATTATATTACTGAACTGTCCGAGAGTTTCCAAATGTAACATATAACTATATCTCA
ACATTATGTAATCACTGTGATGATGCACCTTGTGTTAATGCTTGTCCATTAAACCCTAAGGCGAT
GTATAAGTCAGATAACGGAATAACAATGCATAATCATGAAGCTTGTATCGGTTGTCGTGCTTGTG
AAAAAGCATGTCCATATTCAGTTATTTCTTTAACGAAACAGAGCCATTTGGTGAGTGGAGAAGT
GGAGATGCTGCTGAATTAATAAAAAAGTAGGGGGGGATATTATTCCTTACTCAAATCCAGACAG
AGAAATAACTTATGCAGGAATTAGATCTGCTA**AGAAAGTTGAAAAATGTACTTTCT****GCGATCACA**
GAGTTGCAATGGTGAAAAACCTTACTGTGTGATTCTNTCCTGCTAATCAG

522 bases

MVINLEK**C**AG **CAAC**SI**C**KN ENNVDPGIFW SHYITETVRE FPNVTYNYIS
 TL**C**NH**C**DDAP **CVNAC**PLNPK **AM**YKSDNGIT **M**HNHEA**C**IG**C** R**ACE**KAC**P**YS
 VISFNETEPF GEWRSGDAAE LTKKVGGDII PYSNPDREIT YAGIRS**AK**KV
 EK**CTF****C**DHRV ANGEKPY**C**VI L**S****C*** 174 AA

Figure 23: Above :*OhILAs arrB* gene sequence 522 bases. Below: ArrB protein sequence 174 AA (* included). (*arrB* gene sequence is shown in blue, the RBS is underlined, the cysteine residues are in bold and red, the methionines in cyan, * indicates the stop codon for the ORF, inverted repeats are underlined and highlighted.)

The Fe-sulfur subunit in *OhILAs* is 174 amino acids long and the gene is 522 nucleotides long. The *arrB* gene was the ORF directly downstream of *arrA*. There are 45 intergenic bases between *arrA* and *arrB*. The predicted molecular mass based on protein sequence is 18.9 kDa and the predicted pI is 5.72 (www.expasy.org). The amino acid sequence was analyzed using the signal peptide production program SIGNALP (www.cbs.dtu.dk/services/SignalP/), and similar to all known ArrB subunits from other bacteria the Tat signal was absent. Ribosome binding site occur 23 bp upstream of *arrB* (**GAGG**, fig.23) within the intergenic region of 45 bases between *arrA* and *arrB* genes. An internal, potential transcriptional terminator was identified; inverted repeats are underlined (**AGAAAGTTGAAAAATGTACTTTCT**). This could be a pause site in the DNA sequence that slows down transcription of the gene (Figure 23).

The 4Fe-4S binding domain is conserved and clearly identified in OhILAs ArrB protein sequence, suggesting that this ORF is indeed the Fe-S cluster containing small subunit of the arsenate reductase enzyme in OhILAs. Figure 23 indicates four iron-sulfur clusters at the following cysteines: i) cys8, 11, 14, 18; ii) cys 63, 66, 71, 75; iii) cys 87, 90, 93, 97; and iv) cys 153, 156, 168, 172(173). The first and third iron-sulfur clusters share the binding motif (C-X₂-C-X₂-C-X₃-C) while the second and fourth clusters are different with a (C-X₂-C-X₄-C-X₃-C) and (C-X₂-C-X₁₁-C-X₃-C) motif respectively (Figure 24). The fourth cysteine of the fourth cluster is displaced according to the results from this study. But on close examination of protein sequence alignment with ArrB from *Shewanella* sp. ANA-3, *Wolinella succinogenes* and *B.selenitireducens* strain MLS10 it can be said in conclusion that one amino acid residue is extra (sequencing flaw) at the C-terminus of ArrB and thus the cysteine is one place away from its position in the conserved domain. Further sequencing needs to be done to confirm this (Figure 24).

The protein sequence alignment shows that ArrB, the small Fe-S cluster containing subunit in OhILAs Arr, shares 37% identity and 56% similarity with the ArrB sequences from *B.selenitireducens* MLS10, *W.succinogenes* and *Shewanella* sp. strain ANA-3. *D.hafniense* was removed for better sequence alignments (Figure 24).

CLUSTAL W (1.82) multiple sequence alignment

```

ArrBOhILAs      -----MVINLEKCAGCAACSIACKNENNVPDGIFWSHYITETVREFPNVTYNYISTLC 53
MLS10ArrB       MAKKNYAMTIDLQACIGCAGCAVTCKNENNTDDGVNWMHYTKKESGQFPNIRYDFMPTLC 60
ArrBShew        ---MRLGMVIDLQKCVGCGGCSLACKTENNTNDGIHWSHHIATTEGTFPDVKYTYIPTLC 57
WolArrB         -MAKNYAMVIDLNRVCGCGACDLACKSENNTPVGIHWSNHVVETTGVFPNVGFTYRPTLC 59
                  *.*: *: * **..* :*:*.***. *: * ::      **: : : : .***

ArrBOhILAs      NHCDDAPCVNACPLNPKAMYKSDN-GITMHNHEACIGCRACEKACPYSVISFNETEPFGE 112
MLS10ArrB       NHCDHAPCVMACPVT--AMYKDED-GLTLHDADRCIGCKACMTACPYGVISFNKKDPHQY 117
ArrBShew        NHCDDAPCVKVCPTG--AMHKDKR-GLTLQNNDECIGCKCMNACPYGVISFNAATPHRR 114
WolArrB         NHCDDAPCVRACPTG--AMYKDKEGSMTLHNADKIGCKSCMLADPYHVIYNSKQPHYK 117
                  ****.***. **. **:*. . *:*: : : ****: * * ** * :* *.

ArrBOhILAs      WR-----SGDAAELTKKVGDDIIPYSNPDPREITYAGIRSAKKVEKCTFCDHRVANG 163
MLS10ArrB       WNENNAWSE-LSATPKEVQESAGAD-LPYNPERDYNIEAIRYRGIVEKQCFCDHRLARD 175
ArrBShew        WQDDSEVVANGTVSPLMLLKRTGATATPNENPERGDTYPMIRPKRTTEKCTFCDHRLDKG 174
WolArrB         WTTTKPTVM-GMFSPEKEMADKSGAP-YPYFNVERGKTYEALRFKGLVEKCSFCDHRVKAG 175
                  *      . : . *. * * :* .* :* .*** *****: .

ArrBOhILAs      EKPVCVILSC----- 173
MLS10ArrB       EQPYCVERCPSEAMYVGDLNDPDDKIHELLK-YSHEGLKEELGTPKPKVFYLRKFSK---- 230
ArrBShew        LNPAQVDAQCPSEARVIGDLDDPQSKVSQLIKHLKPMQLKPEAGTGPRVFYIRSFQVKTAY 234
WolArrB         LMPHCVVSCPGDARIFGDLNDPKSSVSQLLAQYPAETLLPHKGTAKVFYIRSYARA--- 232
                  * ** .

```

Figure 24: OhILAs ArrB protein alignment with ArrB from MLS10, *Wolinella succinogenes* and *Shewanella* sp. ANA-3 ([*] indicates identity and [. & :] indicates similarity, all conserved cysteine residues are shown red)

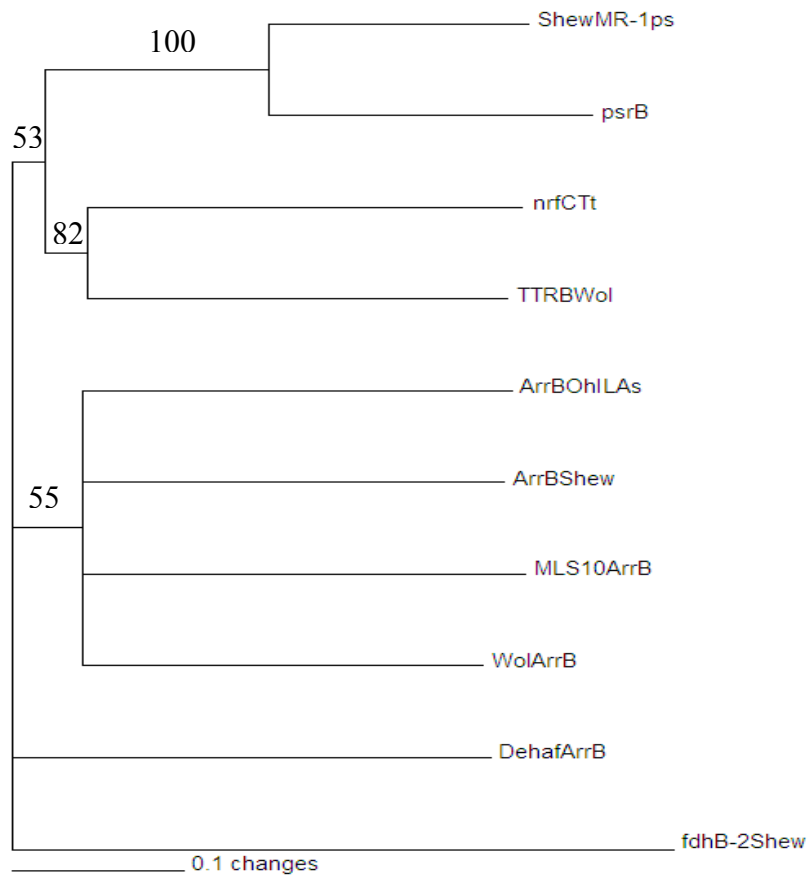


Figure 25: Rooted, neighbor joining tree with bootstrap values generated for protein sequences that showed significant alignments with OhILAs ArrB.

The rooted (*fdhB-Shewanella*), neighbor joining tree with bootstrap values generated for the ArrB sequences that showed significant alignment with OhILAs ArrB, suggests that OhILAs ArrB is similar to the ArrB protein from MLS10, *Shewanella* ANA-3 and *W.succinogenes* (Figure 25).

OhILAs ArrB		
	No.	%
Non polar		
Ala (A)	15	8.7
Val (V)	10	5.8
Leu (L)	5	2.9
Ile (I)	13	7.5
Pro (P)	10	5.8
Met (M)	3	1.7
Phe (F)	5	2.9
Trp (W)	2	1.2
Total	63	36.5
Polar		
Gly (G)	10	5.8
Ser (S)	10	5.8
Thr (T)	9	5.2
Cys (C)	16	9.2
Tyr (Y)	8	4.6
Asn (N)	14	8.1
Gln (Q)	0	0
Total	67	38.7
Acidic		
Asp (D)	8	4.6
Glu (E)	13	7.5
Total	21	12.1
Basic		
Lys (K)	11	6.4
Arg (R)	6	3.5
His (H)	5	2.9
Total	22	12.8
Missing residues	0	0
Total residues	173	
Molecular weight	19 kDa	
Estimated pI	5.72	

Table 7: Amino acid composition of the predicted product of OhILAs *arrB* gene.

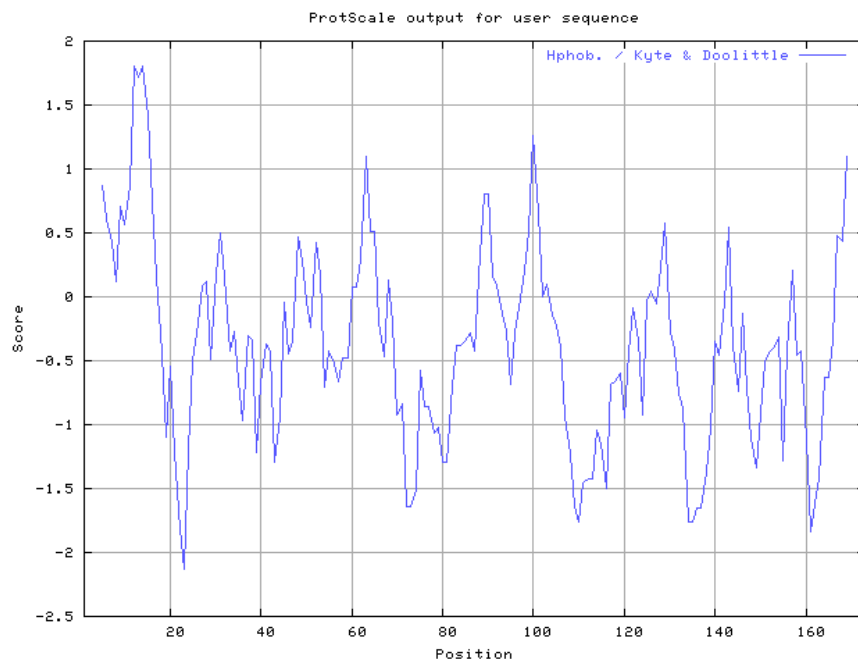


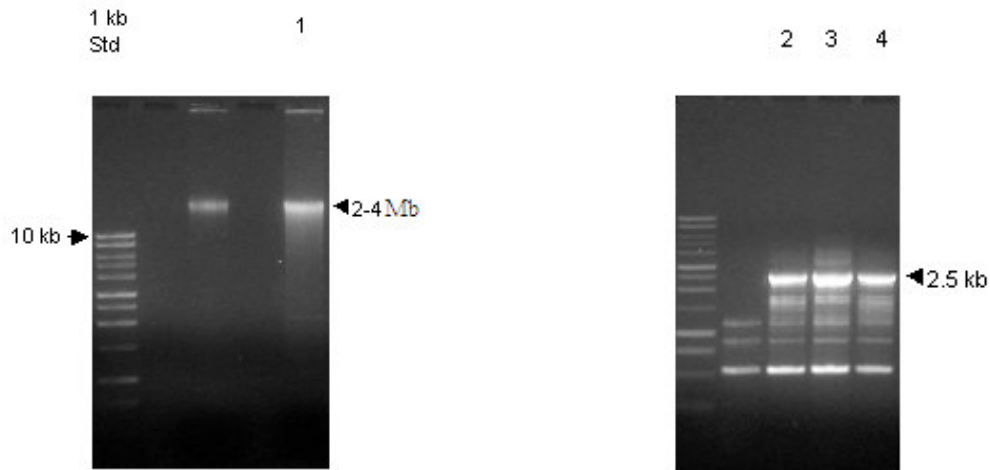
Figure 26: Hydropathy profile of the OhILAs ArrB protein. The average hydropathic index was calculated with an average window size of 9 amino acids. The amino acid position is along the x-axis and the hydropathy score is along the y-axis.

The hydropathy profile based on the amino acid composition of 36.5% non-polar residues and 38.7% polar residues for OhILAs ArrB (Table 7, reveals a more hydrophilic protein (Figure 26). The N terminus is composed of 20 residues that are hydrophobic and the C terminus of another 20 residues that are hydrophilic. The residues 20 – 160 are primarily more hydrophilic than hydrophobic in nature.

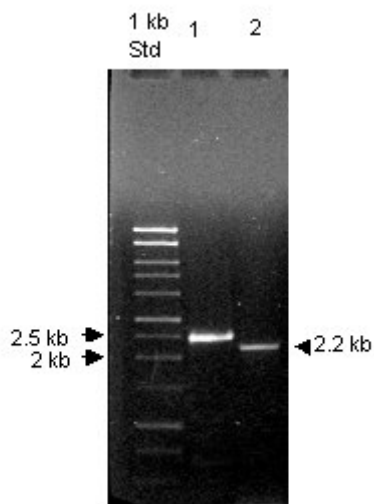
From earlier studies it has been seen that OhILAs arsenate reductase Arr was composed of two subunits which revealed as two bands on activity gels of pure Arr after gel filtration chromatography. The Arr antibody probe based on a 15 amino acid conserved sequence at the C terminus of ArrA from MLS10 recognized the active enzyme from OhILAs on western blots. These results suggested that OhILAs arsenate reductase is also a heterodimer composed of a bigger catalytic subunit ArrA and a smaller

subunit ArrB similar to that from MLS10 or *Shewanella* ANA-3 (Mirunalni Thangavelu, Masters thesis, 2004).

Panel A



Panel B



Panel C

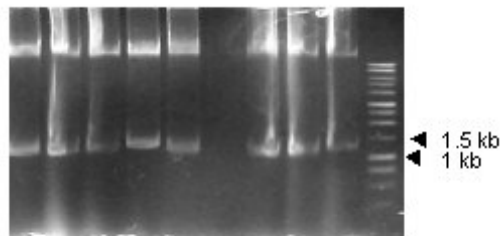
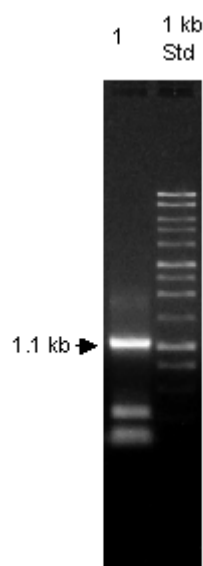


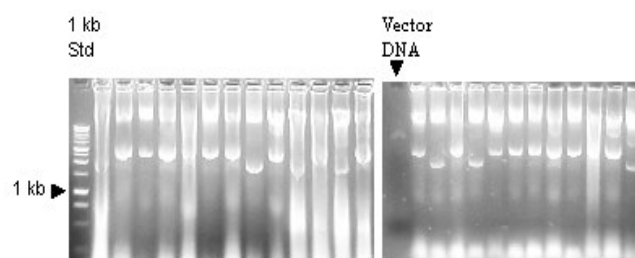
Figure 27: MLS10 *arrA* amplification; *arrB* amplification and cloning.

Panel A: MLS10 *arrA* amplification. Lane 1: MLS10 genomic DNA (~2-4 Mb). Lanes 2-4: PCR product amplified with primers 110JSF1/34JSR2 (~2.5 kb). **Panel B:** MLS10 *arrB* amplification. Lanes 1 and 2: PCR product amplified with primers 34JSF3/ArrBDhR3 (~2.5 kb) and 34JSF4/ArrBShR2 (~2.2 kb). **Panel C:** Cracking results of the cloned 2.5 kb amplicon containing MLS10 *arrB*.

Panel A



Panel B



Panel C

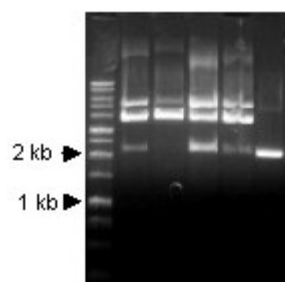
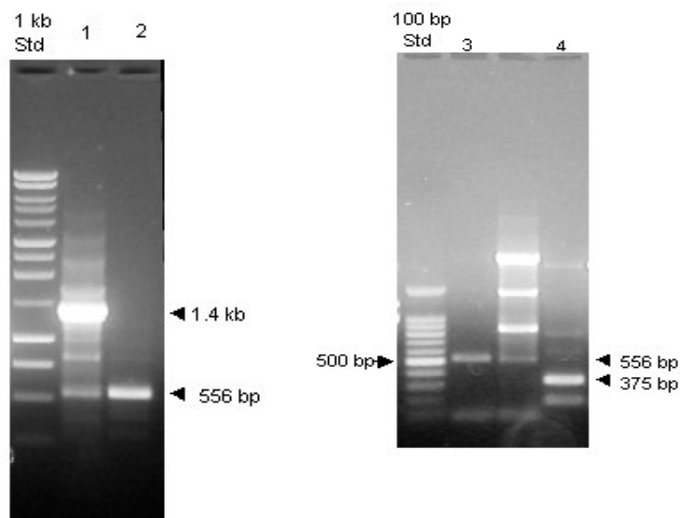


Figure 28: MLS10 *arrB* amplification and cloning.

Panel A: MLS10 *arrB* amplification. Lane 1: PCR product amplified with primers 34JSF4/ArrBDhR3 (~1.1 kb). **Panel B:** Cloning results of the cloned 1.1 kb amplicon containing MLS10 *arrB*. **Panel C:** Restriction digests with EcoR1 of selected clones.

Panel A



Panel B

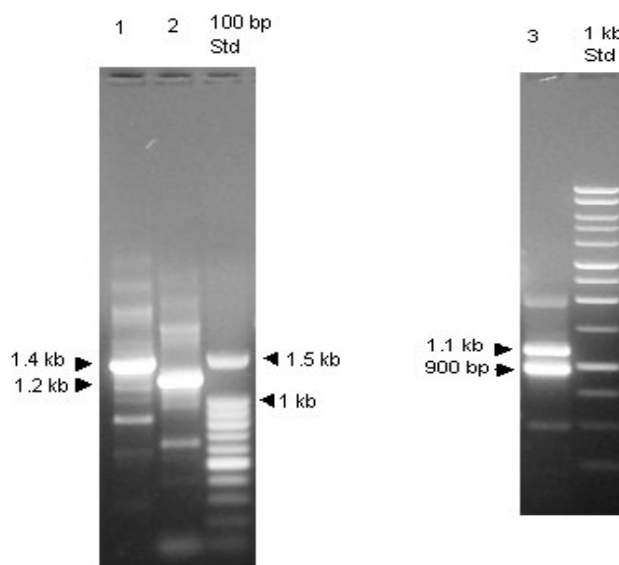


Figure 29: Amplifying the *arrB* gene using newly designed primers.

Panel A: PCR product amplified with: primers MArrBF2/DArRBR3 (~1.4 kb) – Lane 1, primers MArrBF2/DArRBR1 (~556 bp) – Lanes 2 and 3, primers MArrBF3/DArRBR1 (~375 bp) – Lane 4.

Panel B: Lanes 1 and 2: PCR product amplified with primers MArrBF2/DArRBR3 (~1.4 kb) and MArrBF3/DArRBR3 (~1.2 kb), both dominant and expected fragments are seen, arrows indicate the dominant bands. Lane 3: PCR product amplified with primers MArrBF3/DArRBR4, arrows indicate two dominant bands.

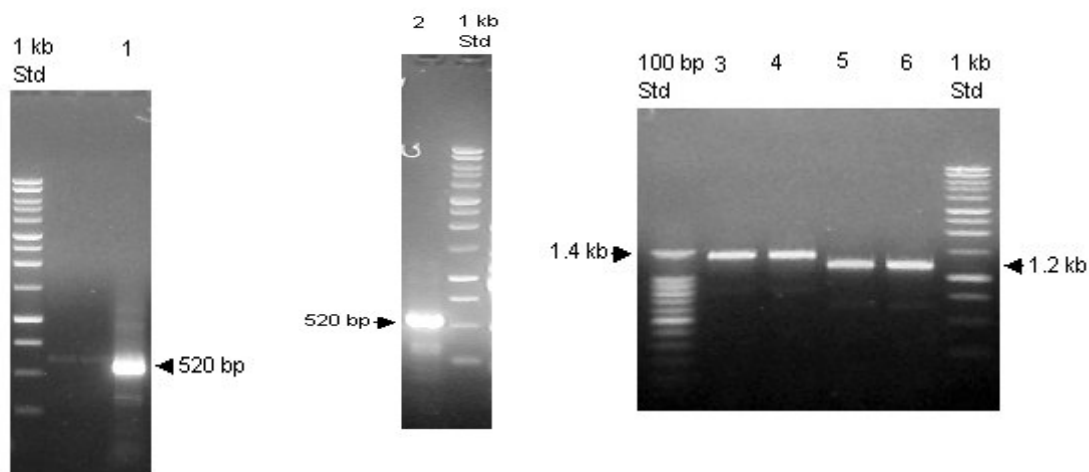
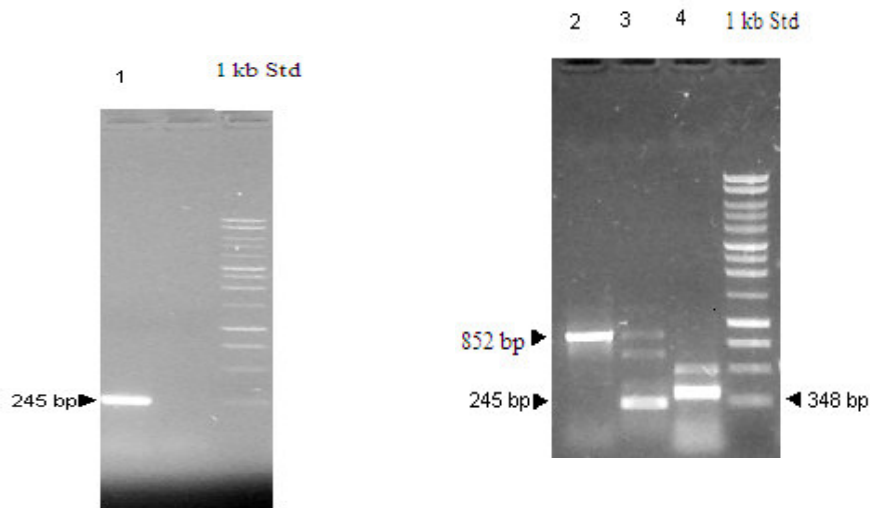


Figure 30: PCR products amplified for sequencing *MLS10 arrB* gene.

Lane 1: PCR product amplified with primers MArrBF3/DArrBR2 (~520 bp) using a 1.4 kb fragment as DNA template amplified with primers MArrBF2/DArrBR3. Lane 2: PCR product amplified with primers MArrBF3/DArrBR2 (~520 bp). Lanes 3 and 4: PCR product (~1.4 kb, MArrBF2/DArrBR3) and Lanes 5 and 6: PCR product (~1.2 kb, MArrBF3/DArrBR3) amplified for elution and purification.

Panel A



Panel B

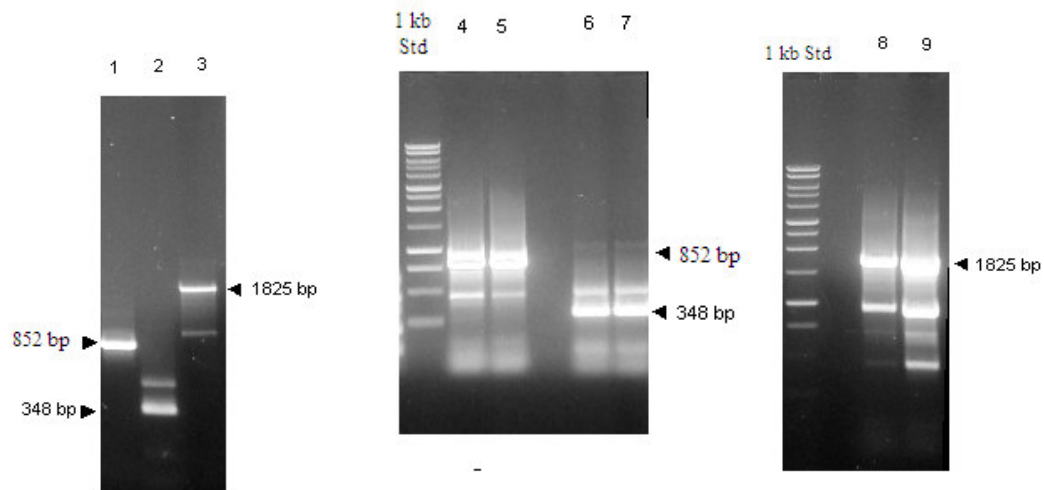


Figure 31: OhILAs *arrA* gene amplification and sequencing.

Panel A: PCR product amplified with: primers arrAUF1/arrAUR2 (~245 bp) – Lanes 1 and 3, primers arrAUF1/arrAUR3 (~852 bp) – Lane 2, primers arrAMLSF4/arrAUR6 (~348 bp) – Lane 4.

Panel B: PCR products amplified for sequencing OhILAs *arrA* gene. Lanes 1, 4 and 5: PCR product amplified with primers arrAUF1/arrAUR3 (~852 bp). Lanes 2, 6 and 7: PCR product amplified with primers arrAMLSF4/arrAUR6 (~348 bp). Lanes 3, 8 and 9: PCR product amplified with primers arrAUF1/arrAUR5 (~1825 bp).

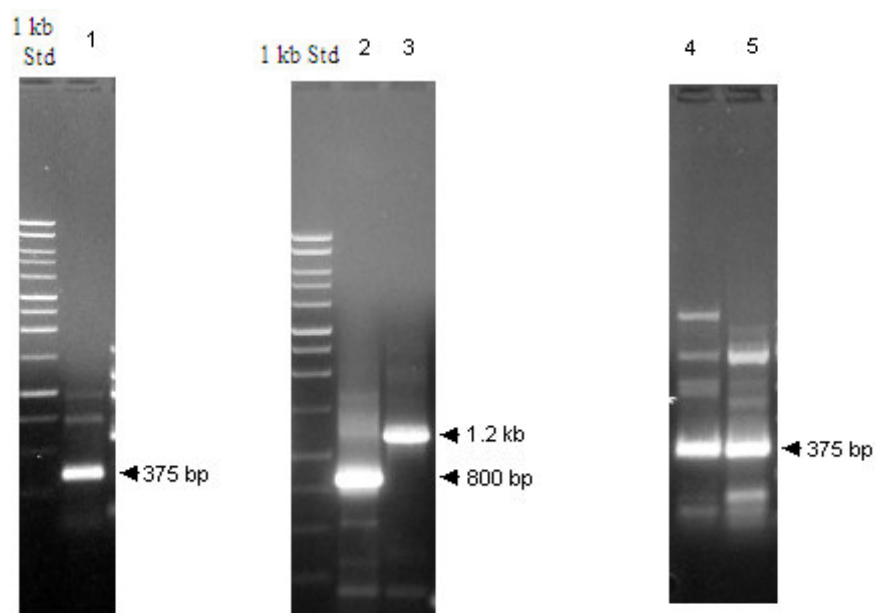


Figure 32: OhILAs *arrB* gene amplification and sequencing.

Lanes 1,4 and 5: PCR product amplified with primers MArrBF3/DArrBR1 (~375 bp).

Lane 2: PCR product amplified with primers arrAUF5/DArrBR4 (~800 bp). Lane 3: PCR product amplified with primers arrAUF5/DArrBR2 (~1.2 kb).

V. DISCUSSION

To date, only two Arr's have been purified and characterized (Krafft and Macy, 1998; Afkar et al., 2003). These include *Chrysiogenes arsenatis* and *Bacillus selenitireducens* strain MLS10. The *arr* operon in *Shewanella* sp. strain ANA-3 was genetically identified and the two ORFs corresponding to *arrA* and *arrB* genes were sequenced (Saltikov and Newman, 2003).

The results from this current study will add to this small list, three new gene sequences pertaining to the arsenate reductases from two low G+C Gram-positive bacteria. It is significant to note that *Shewanella* sp. strain ANA-3 is a member belonging to the γ -Proteobacteria while *Chrysiogenes arsenatis* is the sole member of a deep branch in the bacteria (Oremland and Stolz, 2003; Oremland and Stolz, 2005).

This is the second report on sequence data for Arr after prior studies in *Bacillus selenitireducens* strain MLS10 (Afkari et al., 2003) where the *arrA* gene had been cloned and sequenced and N-terminal data from both ArrA and ArrB was also available. *arrB* gene, however, had not been completely sequenced. The Arr from *D. hafniense* and *W. succinogenes* are inferred proteins.

The current study includes the entire *arrB* gene (693 bases) from MLS10, which was sequenced from genomic DNA, and therefore the amino acid sequence (231 residues) is also now available. Upstream (an intergenic region comprising 29 bases) and downstream sequences (180 bases) are also available.

The gene that encodes the ArrB protein is predicted to be similar to other proteins involved in respiratory reduction such as molybdopterin oxidoreductase, formate

dehydrogenase, arsenate reductases, nitrite reductases etc. ArrB contains iron-sulfur clusters. A Tat motif is absent at the N-terminus as is the case with *Shewanella* ANA-3 and *D.hafniense*. A Ribosome binding site was located 13bp upstream of *arrA* similar to *Shewanella* ANA-3.

The protein identity and similarity, the size of the subunits and the presence of the motifs peculiar to the respiratory arsenate reductases (Arrs) confirms that the *arrB* gene that was sequenced from *B.selenitireducens* strain MLS10 is indeed the smaller subunit of the arsenate reductase, specifically the iron-sulfur (4Fe-4S) containing subunit with a ribosome binding site 13bp upstream of its open reading frame and lying immediately downstream to the *arrA* gene. It is closely related to the other known prokaryotes that use an arsenate reductase for As(V) respiration.

The hydropathy profile of ArrB suggests that it has an equal number of non polar and polar residues, the latter providing the protein a hydrophilic outer surface when it is transported along with ArrA as a complex to the cytoplasmic membrane where it becomes membrane bound. From current understanding of the respiratory arsenate reductase from MLS10 it has been suggested that the enzyme is membrane bound and faces the outside of the cell as it is involved in the respiration of arsenate. This would enable the enzyme to reduce the arsenate to arsenite outside the cell without the need for arsenate entry into the cell, its reduction into the more toxic mobile form i.e. arsenite and therefore necessitating the presence of an As(III) specific efflux protein to rid the cell of this toxic substance.

The presence of cofactors such as molybdenum in ArrA and electron carriers such as the Fe-S cluster in ArrB helps in efficient electron transport through the electron

transport chain aiding in respiration through redox reactions. The oxidation of organic substrates such as lactate and fumarate (electron donors) coupled to the reduction of arsenate (electron acceptor) generates the proton motive force or the proton gradient necessary for energy generation through respiration resulting in growth of the bacterium.

Clostridium sp. strain OhILAs, also a low G+C Gram-positive bacterium was sequenced and two ORFs (+1, +2) corresponding to a partially complete *arrA* gene and an immediately downstream ORF corresponding to *arrB* gene were identified. OhILAs *arrA* gene is ~ 2298 bp and the protein encoded is ~766 residues in length. The *arrB* gene is smaller and a 522 bp in length, the protein it encodes is 174 amino acids in length.

The larger subunit begins with the conserved cysteine motif, but the first two cysteines are missing and need further sequencing upstream. Also, ~122 bases are missing internally, thus, further sequencing is indispensable. The conserved motifs including the molybdopterin dinucleotide binding subunit and the conserved sequence GQGHWA in the C-terminus are present. The protein sequence is predicted to be similar to arsenate, DMSO, polysulfide (*psrA*), thiosulfate and nitrate reductases from BLAST searches. Further sequencing of the entire gene is necessary to ascertain whether a Tat signal sequence is present at the N-terminus.

The smaller subunit is located immediately downstream of the *arrA* gene after an intergenic region comprising 45 bases. No Tat sequence was identified, corroborating findings from other arsenate reductases including MLS10. A Ribosome binding site is present 23 bp upstream of *arrB* gene. The conserved cysteine residues are also present spanning the entire protein sequence in the correct positions confirming the identity of

this protein to be the ArrB , 4Fe-4S cluster containing smaller subunit of the arsenate reductase of OhILAs. An inverted repeat sequence, which could be a potential transcriptional terminator, is internally present in the *arrB* gene sequence at the 3' end. As hypothesized OhILAs arsenate reductase is not very similar to *D.hafniense*, rather it is closely related to *C.arsenatis* for its ArrA protein and to the other known prokaryotes for its ArrB protein (Figure 5).

The hydropathy profile of the two ORFs corresponding to the *arrA* gene suggests a more hydrophilic protein corresponding to the predicted gene product. The evidence is very strong for the larger subunit of the respiratory arsenate reductase Arr, namely ArrA as being composed of a single subunit from *Shewanella* ANA-3, MLS10, *C.arsenatis*, *D. hafniense* and *W. succinogenes* to mention a few characterized and sequenced arsenate reductases. It is therefore too early to conclude that OhILAs ArrA is by itself a two component system composed of two different genes. With the OhILAs ArrB revealing the conserved cysteine motif that could co-ordinate a 4Fe-4S cluster, the Arr from OhILAs is definitely a member of the DMSO reductases family of mononuclear molybdoenzymes. Further sequencing of the *arrA* gene will throw light on this discrepancy of the gene structure of *arrA* from *Clostridium* sp. strain OhILAs, whether it is similar to the larger catalytic subunit ArrA (*arrA*) from MLS10, ANA-3 and *C.arsenatis* or if indeed it has a novel gene order for the arsenate reductase operon *arr*, specific to itself.

The arsenate reductases appear to define a separate class of reductases that are probably specific only for As(V)(Saltikov and Newman, 2003) and are very different from the detoxifying (Ars) arsenate reductases. The Arr are specific for respiration and

are used for the generation of energy. The Arr are closely related to the arsenite oxidase Aox, with the larger subunits having the conserved cysteine motif and the smaller subunit containing a 2Fe-2S binding cysteine motif, the Reiske subunit, which places the latter in the DMSO reductases family of molybdoenzymes along with the former class, the arsenate reductases. The electron flow in the Aox is the reverse of that in the Arr, which is similar to a classic bioenergetics chain for anaerobic respiration.

Phylogenetically the arsenate reductase genes appear to have a broad phylogenetic distribution comprising of members of the Gram-positive, γ , δ , and ϵ –Proteobacteria as well as the archaea from recent study (Figure 4) (Oremland and Stolz, 2005).

VI. FUTURE DIRECTIONS

The upstream of MLS10 *arrA* gene is yet to be sequenced. This will give information regarding the Tat signal sequence for *arrA* helping to classify the encoded protein as a secretory/non-secretory protein. The sequence data will provide further information regarding the presence/absence of another gene (*arrC*) upstream of *arrA* similar to the gene structure in *W. succinogenes* and *D. hafniense*.

The 5' and upstream sequence of OhILAs *arrA* needs to be sequenced for reasons similar to that for MLS10. The *arrA* gene is not completely sequenced and has ~122 bases missing internally. These 122 bases along with the 3' of the *arrB* gene need to be sequenced for better analysis of the arsenate reductase from *Clostridium* sp. strain OhILAs.

VII. REFERENCES:

1. Afkar, E., Lisak, J., Saltikov, C., Basu, P., Oremland, R.S. and Stolz, J.F. 2003. The respiratory arsenate reductase from *Bacillus selenitireducens* strain MLS10.FEMS Microbiol Lett 226 (1): 107-12.
2. Ahmann, D., Roberts, A.L., Krumholz, L.R. and Morel, F.M. 1994. Microbe grows by reducing arsenic.Nature 371 (6500): 750.
3. Anawar, H.M., Akai, J., Mostofa, K.M., Safiullah, S. and Tareq, S.M. 2002. Arsenic poisoning in groundwater: health risk and geochemical sources in Bangladesh.EnvIRON Int 27 (7): 597-604.
4. Anderson, G.L., Williams, J. and Hille, R. 1992. The purification and characterization of arsenite oxidase from *Alcaligenes faecalis*, a molybdenum-containing hydroxylase.J Biol Chem 267 (33): 23674-82.
5. Ayotte, J.D., Montgomery, D.L., Flanagan, S.M. and Robinson, K.W. 2003. Arsenic in groundwater in eastern New England: occurrence, controls, and human health implications.EnvIRON Sci Technol 37 (10): 2075-83.
6. Bennett, M.S., Guan, Z., Laurberg, M. and Su, X.D. 2001. *Bacillus subtilis* arsenate reductase is structurally and functionally similar to low molecular weight protein tyrosine phosphatases.Proc Natl Acad Sci U S A 98 (24): 13577-82.
7. Berks, B.C. 1996. A common export pathway for proteins binding complex redox cofactors?Mol Microbiol 22 (3): 393-404.
8. Carlin, A., Shi, W., Dey, S. and Rosen, B.P. 1995. The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance.J Bacteriol 177 (4): 981-6.
9. Cervantes, C., Ji, G., Ramirez, J.L. and Silver, S. 1994. Resistance to arsenic compounds in microorganisms.FEMS Microbiol Rev 15 (4): 355-67.
10. Christen, K. 2001. Chickens, manure, and arsenic.EnvIRON Sci Technol 35 (9): 184A-185A.
11. Croal, L.R., Gralnick, J.A., Malasarn, D. and Newman, D.K. 2004. The genetics of geochemistry.Annu Rev Genet 38 175-202.

12. Cullen, W.R.a.R., K.J. 1989. Arsenic speciation in the environment. *Chem. Rev.* 89 713-764.
13. Ellis, P.J., Conrads, T., Hille, R. and Kuhn, P. 2001. Crystal structure of the 100 kDa arsenite oxidase from *Alcaligenes faecalis* in two crystal forms at 1.64 Å and 2.03 Å. *Structure (Camb)* 9 (2): 125-32.
14. Garbarino, J.R., Bednar, A.J., Rutherford, D.W., Beyer, R.S. and Wershaw, R.L. 2003. Environmental fate of roxarsone in poultry litter. I. Degradation of roxarsone during composting. *Environ Sci Technol* 37 (8): 1509-14.
15. Gihring, T.M. and Banfield, J.F. 2001. Arsenite oxidation and arsenate respiration by a new *Thermus* isolate. *FEMS Microbiol Lett* 204 (2): 335-40.
16. Gihring, T.M., Druschel, G.K., McCleskey, R.B., Hamers, R.J. and Banfield, J.F. 2001. Rapid arsenite oxidation by *Thermus aquaticus* and *Thermus thermophilus*: field and laboratory investigations. *Environ Sci Technol* 35 (19): 3857-62.
17. Gladysheva, T.B., Oden, K.L. and Rosen, B.P. 1994. Properties of the arsenate reductase of plasmid R773. *Biochemistry* 33 (23): 7288-93.
18. Hindmarsh, J.T. 2000. Arsenic, its Clinical and Environmental Significance. *The Journal of Trace Elements in Experimental Medicine* 13: 165-172.
19. Krafft, T. and Macy, J.M. 1998. Purification and characterization of the respiratory arsenate reductase of *Chrysiogenes arsenatis*. *Eur J Biochem* 255 (3): 647-53.
20. Krafft, T., Bowen, A., Theis, F. and Macy, J.M. 2000. Cloning and sequencing of the genes encoding the periplasmic-cytochrome B-containing selenate reductase of *Thauera selenatis*. *DNA Seq* 10 (6): 365-77.
21. Macy, J.M., Nunan, K., Hagen, K.D., Dixon, D.R., Harbour, P.J., Cahill, M. and Sly, L.I. 1996. *Chrysiogenes arsenatis* gen. nov., sp. nov., a new arsenate-respiring bacterium isolated from gold mine wastewater. *Int J Syst Bacteriol* 46 (4): 1153-7.
22. Marchler-Bauer, A. and Bryant, S.H. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* 32 (Web Server issue): W327-31.
23. Martin, P., DeMel, S., Shi, J., Gladysheva, T., Gatti, D.L., Rosen, B.P. and Edwards, B.F. 2001. Insights into the structure, solvation, and mechanism of ArsC arsenate reductase, a novel arsenic detoxification enzyme. *Structure (Camb)* 9 (11): 1071-81.

24. Mukhopadhyay, R. and Rosen, B.P. 1998. *Saccharomyces cerevisiae* ACR2 gene encodes an arsenate reductase.FEMS Microbiol Lett 168 (1): 127-36.
25. Mukhopadhyay, R. and Rosen, B.P. 2002. Arsenate reductases in prokaryotes and eukaryotes.Environ Health Perspect 110 Suppl 5 745-8.
26. Mukhopadhyay, R., Li, J., Bhattacharjee, H. and Rosen, B.P. 1998. Metalloid resistance mechanisms.Adv Exp Med Biol 456 159-81.
27. Mukhopadhyay, R., Rosen, B.P., Phung, L.T. and Silver, S. 2002. Microbial arsenic: from geocycles to genes and enzymes.FEMS Microbiol Rev 26 (3): 311-25.
28. Newman, D.K., Kennedy, E.K., Coates, J.D., Ahmann, D., Ellis, D.J., Lovley, D.R. and Morel, F.M. 1997. Dissimilatory arsenate and sulfate reduction in *Desulfotomaculum auripigmentum* sp. nov.Arch Microbiol 168 (5): 380-8.
29. Nriagu, J.O.e. 1994. Arsenic in the environment.Advances in environmental science and technology 27
30. Oremland, R.S. and Stolz, J.F. 2003. The ecology of arsenic.Science 300 (5621): 939-44.
31. Oremland, R.S. and Stolz, J.F. 2005. Arsenic, microbes and contaminated aquifers.Trends Microbiol 13 (2): 45-9.
- 32.Oremland, R.S., Stolz, J.F. and Hollibaugh, J.T. (2004). The Microbial Arsenic Cycle in Mono Lake California. Microbiology: 15-27.
33. Oremland, R.S., Hoefft, S.E., Santini, J.M., Bano, N., Hollibaugh, R.A. and Hollibaugh, J.T. 2002. Anaerobic oxidation of arsenite in Mono Lake water and by a facultative, arsenite-oxidizing chemoautotroph, strain MLHE-1.Appl Environ Microbiol 68 (10): 4795-802.
34. Oremland, R.S., Herbel, M.J., Blum, J.S., Langley, S., Beveridge, T.J., Ajayan, P.M., Sutto, T., Ellis, A.V. and Curran, S. 2004. Structural and spectral features of selenium nanospheres produced by Se-respiring bacteria.Appl Environ Microbiol 70 (1): 52-60.
35. Osborne, F.H. and Enrlich, H.L. 1976. Oxidation of arsenite by a soil isolate of *Alcaligenes*.J Appl Bacteriol 41 (2): 295-305.
36. Rosen, B.P. 2002. Biochemistry of arsenic detoxification.FEBS Lett 529 (1): 86-92.
37. Ryker, S.J. 2001. Mapping arsenic in groundwater.Geotimes 46: ((11)): 34-36.

38. Saltikov, C.W. and Newman, D.K. 2003. Genetic identification of a respiratory arsenate reductase. *Proc Natl Acad Sci U S A* 100 (19): 10983-8.
39. Saltikov, C.W., Cifuentes, A., Venkateswaran, K. and Newman, D.K. 2003. The ars detoxification system is advantageous but not required for As(V) respiration by the genetically tractable *Shewanella* species strain ANA-3. *Appl Environ Microbiol* 69 (5): 2800-9.
40. Silver, S. and Phung, L.T. 2005. Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Appl Environ Microbiol* 71 (2): 599-608.
41. Stolz, J.F. and Oremland, R.S. 1999. Bacterial respiration of arsenic and selenium. *FEMS Microbiol Rev* 23 (5): 615-27.
42. Stolz, J.F., Ellis, D.J., Blum, J.S., Ahmann, D., Lovley, D.R. and Oremland, R.S. 1999. *Sulfurospirillum barnesii* sp. nov. and *Sulfurospirillum arsenophilum* sp. nov., new members of the *Sulfurospirillum* clade of the epsilon Proteobacteria. *Int J Syst Bacteriol* 49 Pt 3 1177-80.
43. Switzer Blum, J., Burns Bindi, A., Buzzelli, J., Stolz, J.F. and Oremland, R.S. 1998. *Bacillus arsenicoselenatis*, sp. nov., and *Bacillus selenitireducens*, sp. nov.: two haloalkaliphiles from Mono Lake, California that respire oxyanions of selenium and arsenic. *Arch Microbiol* 171 (1): 19-30.
44. vanden Hoven, R.N. and Santini, J.M. 2004. Arsenite oxidation by the heterotroph *Hydrogenophaga* sp. str. NT-14: the arsenite oxidase and its physiological electron acceptor. *Biochim Biophys Acta* 1656 (2-3): 148-55.

VIII. Appendices

Appendix 1: BASAL SALT MEDIUM for *Bacillus selenitireducens* strain MLS10. Lactate modified medium from the recipe of Switzer Blum (Switzer Blum et al., 1998) supplemented with vitamins and trace elements (15ml/l) as described in Oremland et al., 1999.

CONSTITUENTS	g/L
Na ₂ CO ₃	10.6
NaHCO ₃	4.2
(NH ₄) ₂ SO ₄	0.1
MgSO ₄ ·7H ₂ O	0.025
NaCl	90
K ₂ HPO ₄	0.15
KH ₂ PO ₄	0.08
Cysteine-HCl	0.125
Na ₂ WO ₄	9x10 ⁻⁶
Yeast extract	0.2
Sodium Arsenate (10mM)	3.12
Additions	
Sodium Lactate	2.1ml*
Trace elements	15ml †
Vitamin mix	15ml †

* 2.1mls provided from 60% Sodium Lactate solution (Sigma Aldrich Co., St. Louis , MO, USA) to provide a final concentration of 15mM in the medium

† Vitamins and Trace elements are provided to the medium after autoclaving from sterile stock solutions of concentrations (100X) each.

Mineral Mix	g/L
Nitrilotriacetic Acid	1.5
Magnesium Sulfate	3.0
Manganese Chloride	0.44
Sodium Chloride	1.0
Ferric Chloride	0.067
Calcium Chloride	0.1
Cobalt Chloride	0.1
Zinc Sulfate	0.27
Copper Sulfate	0.01
Aluminium Potassium Sulfate	0.001
Boric Acid	0.01
Sodium Molybdate	0.025
Nickel Chloride	0.02
Sodium Tungstate	0.025

Vitamin Mix	mg/L
Biotin	2
Folic Acid	2
Pyridoxine hydrochloride	10
Thiamine	5
Nicotinic Acid	5
Pantothenic Acid	5
p-Aminobenzoic acid	5
Thioctoc Acid	5
Vitamin B12	0.1

Appendix 2: BASAL SALT MEDIUM for *Clostridium* sp. OhILAs

CONSTITUENTS	g/L
MgCl	0.095gm
NaHCO ₃	4.2gm
K ₂ HPO ₄	0.225gm
KH ₂ PO ₄	0.225gm
NaCl	1.0gm
NH ₄ Cl	0.178gm
Sodium Sulfide	0.1gm
Yeast	1 gm
Trace elements (100x)	10ml
Vitamin mix (100x)	10ml
<u>Electron donor</u>	
Sodium Lactate (2.8ml of 60%w/w sodium lactate solution)	2.24gm
Na Acetate (20mM)	2.72gm
Pyruvate (10mM)	1.1gm
Formate (20mM)	1.36gm
Glycerol	1.1842gm
<u>Electron acceptor</u>	
Sodium Arsenate (20mM)	6gm

Appendix 3: Genomic DNA extraction protocol

This protocol can be used for DNA extraction from - **MLS10, MIT-13, SES-3, *S.del*, JMM4 and OhILAs**

For a 5-10ml sample of cells (amounts may be scaled up):

1. Spin cells at 7500 x g for 15 minutes to pellet cells.
2. Discard supernatant. Wash cells in half the volume of lysis buffer (50mM EDTA, 0.1M NaCl, pH 7.5)
3. Repeat step 1. Discard supernatant.
4. Resuspend cells in 800µl lysis buffer.
5. Add 90ul lysozyme (0.1 g/ml lysis buffer).
6. Incubate for 10 minutes at 37°C without shaking.
7. Add 60ul 20% SDS and mix.
8. Add 800ul Tris-buffered phenol (1M Tris.Cl pH 8.0 : Liquid phenol, 1:1).
9. Invert 4 times to mix and centrifuge for 5 minutes at 12,000 x g.
10. Withdraw upper aqueous phase carefully from the lower phenolic phase.
11. Precipitate DNA by adding 40µl 5M NaCl and 2ml cold 100% Ethanol. Leave at -20°C for about 30 minutes.
12. Spin sample at 12,000 x g at 4°C for 10 minutes to retrieve DNA.
13. Wash pellet (very tiny, sticking to sides of the centrifuge tube) with cold 70% Ethanol.
14. Dry pellet in speed vacuum at medium heat and RNase treat the genomic DNA.
15. Resuspend in 200µl sterile water/TE buffer according to future protocol needs.
16. Store DNA at -20°C

Appendix 4: RNase treatment of genomic DNA protocol

1. Thaw DNA sample in 400µl sterile water/TE buffer.
2. Add 2µl RNase (10mg/ml in 10mM Tris pH 7.6, 15mM NaCl) to the DNA sample i.e. 1:200,v/v and heat on a 65°C water bath for 10 minutes.
3. Add 100% cold ethanol (2:1), put into -20°C for 30 minutes.
4. Spin at 12,000 rpm for 10 minutes and decant the ethanol.
5. Dry DNA in a speed vacuum. Then resuspend in 200µl TE buffer/sterile water.
6. Run 1µl of DNA on a 0.8% agarose gel to check for single DNA band.

Appendix 5: Direct sequencing protocol (Dr. David Lampe's Lab)

- 1) Purify the template DNA using CENTRI-SEP spin columns. (Refill used columns with 800ul Sephadex G-50 slurry [2g in 32ml water]).
- 2) Quantify the purified template DNA before setting up the sequencing reaction.
- 3) Set up the sequencing reaction

The following is (Dr. Lampe's) standard recipe:

300ng Plasmid DNA (eluted in ddH₂O)

2.0ul Big Dye Reagent

6.0ul 2.5 X sequencing buffer

1.0ul Primer (from 20uM stock)

Water to 20ul

DNA template quantities to use in a cycle sequencing reaction:

PCR product:

100 -200 bp – 1-3 ng

200 -500 bp – 3-10 ng

500 -1000 bp – 5-20 bp

1000 -2000 bp – 10-40 ng

>2000 bp – 40-100 ng

Single stranded plasmid – 50-100 ng

Double stranded plasmid – 200-500 ng

Cosmid, BAC – 0.5-1.0 ng

Bacterial genomic DNA – 2-3 ng

Cycling Parameters:

- | | | |
|-----------------------------|------|------------|
| 1. Initial denaturation | 95°C | 1 minute |
| 2. Denaturation | 95°C | 10 seconds |
| 3. Annealing | 50°C | 5 seconds |
| 4. Extension | 60°C | 4 minutes |
| 5. Steps 2 to 4 – 55 cycles | | |

4) Clean the sequencing reaction again. Repeat step 1.

5) Lyophilize until dry at medium heat in a speedvac (around 30 minutes).

6) Resuspend in 20ul deionized formamide. Vortex and spin down.

7) Heat to 95°C for 2 minutes.

8) Quick cool on ice. Spin down and store at -20°C until put on the sequencer. (formamide degrades DNA at room temperature)

Appendix 6: Bacterial Media – Luria-Bertani (LB Medium), SOC and SOB Medium

LB Agar (per liter)

10 g NaCl

10 g Tryptone

5 g Yeast extract

15 g Agar

Add deionized water to a final volume of 1 liter

Adjust pH to 7.0 with 5N NaOH

Autoclaved

Cool to 55°C

Add 50ug/ml Kanamycin or 100ug/ml Ampicillin

Pour into petri plates

Cool to room temperature

SOB Medium (per liter)

20 g Tryptone

5 g Yeast extract

0.5 g NaCl

Add 10ml of a 250mM solution of KCl

Adjust pH to 7.0 with 5N NaOH

Add deionized water to a final volume of 1 liter

Autoclave

SOC Medium (per liter)

SOB Medium

Add 20ml of a 1M filter-sterilized solution of glucose and filter sterilize.

Appendix 7: Cracking procedure protocol

(This procedure helps in rapid disruption of colonies to test for inserts in plasmids)

1. Inoculate a new LB+antibiotic plate with a small amount of each colony to be used for cracking analysis.
2. Pick individual colonies from transformation plates using sterile toothpick.
3. Smear to the bottom of sterile 1.5ml microfuge tube.
4. Add 50ul of 10mM EDTA, pH 8.0.
5. Add 50ul of freshly made cracking buffer (25ul 10m NaOH, 0.025g SDS, 250ul 100mM EDTA and bring upto 5ml with DDH₂O).
6. Vortex twice.
7. Incubate at 70 °C for 15 minutes. Allow sample to cool to room temperature.
8. Add 1.5ul 4M KCl and 0.5ul 0.4% Bromophenol blue. Vortex.
9. Place on ice for 5 minutes.
10. Spin for 3 minutes at maximum speed.
11. Run 25-35ul of supernatant on 0.8% agarose gel, remembering to run about 0.1ul of vector DNA as control.

(Once a positive transformant is found, cells are grown for plasmid isolation in LB broth [containing antibiotic] overnight.)